Running head: FOUR DIMENSIONS UNDERLIE FACE IMPRESSIONS

Comprehensive trait attributions show that face impressions are organized in four dimensions

Chujun Lin[1*], Umit Keles[1], Ralph Adolphs[1,2].

[1]Division of Humanities and Social Sciences, California Institute of Technology, CA, USA.

[2]Division of Biology and Biological Engineering, California Institute of Technology, CA, USA.

*Correspondence to: clin7@caltech.edu.

Abstract: How do people form impressions of others based on faces? Existing psychological theories argue that people attribute traits to others from faces along two or three dimensions. While these theories have now been incorporated into numerous empirical and theoretical studies, they were derived from a small set of trait attributions, which limits their generalizability and leaves the true nature of the psychological dimensions unclear. The present study applied deep neural networks to representatively sample an inclusive list of traits and faces, generating a comprehensive set of 100 traits and 100 faces that we administered in two large-scale preregistered studies. These comprehensive trait attributions (Study 1, 750,000 ratings) revealed a novel four-dimensional space: warmth, competence, female-stereotype, and youth-stereotype, challenging existing theories. Study 2 collecting dense individual-level data in seven different countries (2,100,000 trials) reproduced this four-dimensional space across cultures and in individual participants. These findings, together with test-retest reliability of all trait attributions and direct comparisons with existing theories, provide a new, most comprehensive characterization of trait attributions from faces.

One Sentence Summary: Two preregistered studies across seven countries revealed that four novel dimensions underlie face impressions, challenging existing theories.

Main Text:

With the widespread availability of both the tools for manipulating face images for social media posting (*1–3*), and automated methods for categorizing emotions and traits from face images (*4–7*), faces remain a ubiquitous and important source of social information. Upon viewing a face, humans spontaneously attribute a wide range of traits to the individual, such as attributions of demographics (e.g., gender, age), physical appearance (e.g., baby-faced, beautiful), social evaluation (e.g., trustworthy, competent), and personality (e.g., aggressive, sociable) (*8–13*). Although the ground-truth diagnostic validity of trait attributions from faces remains inconclusive (*14–22*), they nonetheless have important consequences for social decisions, ranging from decisions of who to trust and who to punish in laboratory experiments (*23–25*) to decisions of who to elect and who to jail in the real world (*26–31*).

Despite the considerable amount of work on the topic (*8–34*), it remains unclear how to characterize the large number and variety of traits that humans attribute to faces. What is the psychological space that underlies these attributions? Pioneering work attempting to address this question asked participants to attribute thirteen traits (mostly personality traits and social evaluative qualities) to unfamiliar faces and showed that those attributions were represented by two dimensions: trustworthiness and dominance (*8*). A subsequent study that analyzed attributions of a distinct set of thirteen traits (with an emphasis on traits about physical appearance) instead revealed a different three-dimensional space: approachability, dominance, and youthful/attractiveness (*35*). The two or three dimensional models from these initial studies have been regarded as the canonical frameworks in the field for the past decade and have now been incorporated into numerous subsequent studies (*33*, *36–50*).

However, all studies so far have focused on a relatively small number of traits (typically 2 to 25 traits) and omitted a substantial number of traits that humans regularly attribute to faces. In part as a consequence, the accumulating discrepant findings on the dimensionality of face impressions (*8*, *9*, *35*, *37*, *38*, *40*, *51*) have been left unresolved. Here, we argue that to characterize the comprehensive space of trait attributions from faces, it is essential to investigate a more inclusive list of traits that cover all relevant categories. This has been missing in the literature due in good part to a major methodological and practical challenge: one would have to adequately sample many traits that span multiple domains such that they are also meaningful and non-redundant, and to collect reliable attributions for all those traits.

We set out to meet this challenge in the present project. We assembled an exhaustive list of adjectives from multiple sources that describe a person's demographic characteristics, physical appearance, social evaluative qualities, personality, and emotional traits, and applied a pre-trained neural network to sample this list and derive a representative and non-redundant subset of 100 final trait stimuli (Fig. 1A-D) [see also supplementary materials and methods section M1]. Similarly, we combined multiple extant face databases and applied a pre-trained neural network to derive a representative and non-redundant subset of 100 final face stimuli (Fig. 1E-H) [see also supplementary materials and methods section M2]. We verified that our 100 traits were representative of the words people freely generate for our face stimuli. Attributions of the 100 traits from the 100 faces (750,000 sparse online ratings from 1,500 participants with test-retest reliability assessed for every trait) revealed a novel four-dimensional space (Study 1). To test the generalizability of this finding and to address the issue of sparse individual-level data, Study 2 collected complete datasets from every participant in seven different countries and regions

(10,000 trials in each of 210 participants across North America, Latvia, Peru, the Philippines, India, Kenya, and Gaza), and largely reproduced this novel four-dimensional space in both aggregated and individual data. We show that this novel four-dimensional framework better captures the variance of most trait attributions than do existing theories. All experiments were preregistered on Open Science Framework (including participants, materials, procedures, analyses, etc.; see http://bit.ly/osfpre1, http://bit.ly/osfpre2, http://bit.ly/osfpre3, and http://bit.ly/osfpre4).

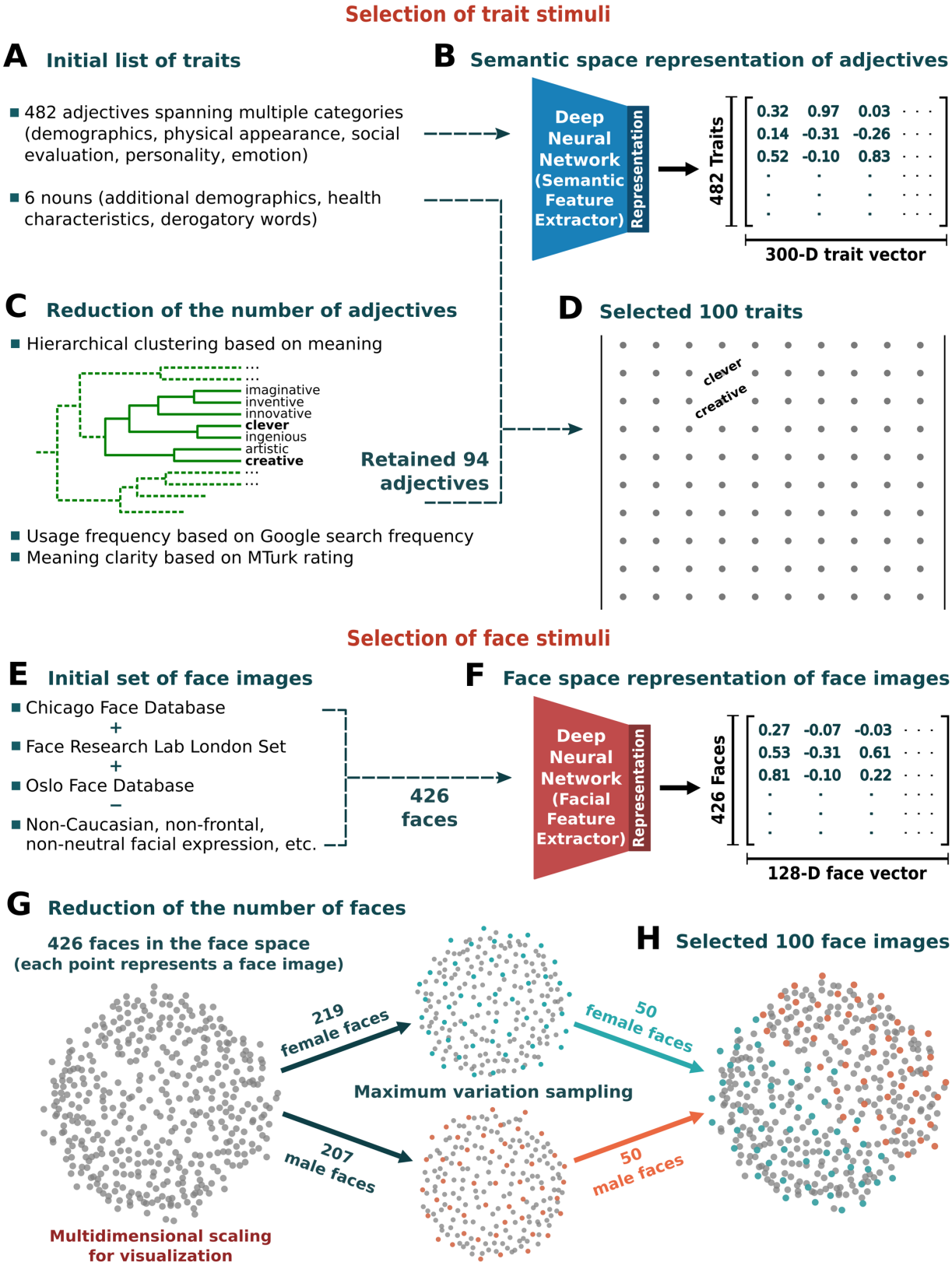

Selection of trait stimuli
A Initial list of traits
482 adjectives spanning multiple categories (demographics, physical appearance, social evaluation, personality, emotion)
6 nouns (additional demographics, health characteristics, derogatory words)
B Semantic space representation of adjectives
Deep Neural Network (Semantic Feature Extractor)
Representation
482 Traits
0.32 0.97 0.03
0.14 -0.31 -0.26
0.52 -0.10 0.83
300-D trait vector
C Reduction of the number of adjectives
Hierarchical clustering based on meaning
imaginative
inventive
innovative
clever
ingenious
artistic
creative
Retained 94 adjectives
Usage frequency based on Google search frequency
Meaning clarity based on MTurk rating
D Selected 100 traits
clever
creative
Selection of face stimuli
E Initial set of face images
Chicago Face Database
+
Face Research Lab London Set
+
Oslo Face Database
−
Non-Caucasian, non-frontal, non-neutral facial expression, etc.
426 faces
F Face space representation of face images
Deep Neural Network (Facial Feature Extractor)
Representation
426 Faces
0.27 -0.07 -0.03
0.53 -0.31 0.61
0.81 -0.10 0.22
128-D face vector
G Reduction of the number of faces
426 faces in the face space
(each point represents a face image)
219 female faces
207 male faces
Maximum variation sampling
Multidimensional scaling for visualization
H Selected 100 face images
50 female faces
50 male faces

Fig. 1 Sampling trait stimuli (A-D) and face stimuli (E-H) to generate a comprehensive set.

(**A**) We gathered an inclusive list of adjectives and nouns that describe a person's demographic characteristics, physical appearance, social evaluative qualities, personality, and emotional traits from multiple sources (*8*, *9*, *15*, *26*, *28*, *32*, *33*, *52–60*). (**B**) Many of the 482 adjectives in (**A**) shared similar or opposite meanings. To quantify the similarity between these adjectives, we represented each of them with a vector of 300 computationally extracted semantic features (that describe word embeddings and text classification) using a state-of-the-art neural network (*61*) that had been pre-trained to assign words to their contexts across 600 billion words. (**C**) Three filters were applied to remove adjectives with similar meanings [as assessed by the cosine distances between the 300-feature word vectors generated in (**B**)], adjectives with unclear meaning, and adjectives with infrequent usage. (**D**) The 94 sampled adjectives together with 6 nouns of additional demographic and health characteristics (education, income, sexual orientation, autism) and derogatory words (idiot, loser) comprised the final trait set (see Table S1 for the complete list of traits and their definitions—a one-sentence definition of each trait was provided to participants in our study to eliminate possible heterogeneity in how each individual understands the meaning of a trait word). (**E**) Aiming to derive a representative set of face images that are of high quality, we first combined three face databases largely used in the literature (*62–64*) to yield 426 Caucasian faces that were clear, frontal, with direct eye gaze and neutral expression, and without glasses or other objects obscuring the face. (**F**) To quantify the similarity in facial structure between the faces gathered in (**E**), we represented each face with a vector of 128 computationally extracted facial features using a state-of-the-art neural network (*65*) that had been pre-trained to identify individuals across millions of faces (of all different aspects and races). (**G**) Maximum variation sampling (*66*) was applied to select faces with maximum variability in facial structure [i.e., maximum dissimilarity, as assessed by the Euclidean distances between the 128-feature face vectors generated in (**E**)], and a final set of 100 faces was obtained (**H**).

Results

A four-dimensional space characterizes trait attributions from faces

To verify that our selected 100 traits were indeed representative of the attributions people spontaneously make from faces, we first collected an independent dataset from participants who freely generated any word about the person that came to mind upon viewing the faces (Fig. S1A). All freely generated words (973 words in total; words that appeared only once were excluded, as they were comprised mainly of misspelled words) were found to be similar to at least one of our selected traits (the similarity between two words was assessed with the cosine distance between the 300-feature vectors of the two words; except for "giving," "moving," and "round", all similarities between freely generated words and their closest counterparts in our trait set > 0.25, which was the mean similarity among our selected 100 traits; 78 freely generated words were identical to those in our 100 selected traits; see Fig. S1B).

For our main preregistered Study 1, we applied exploratory factor analysis (EFA) on aggregate-level data (ratings averaged across participants per trait per face) to analyze the latent structure of the trait attributions. We confirmed that our data were reliable for those analyses: ratings across faces for all traits showed sufficient variance (mean range across traits = 3 points on our 7-point Likert scale; Fig. S2) as well as satisfactory within-subject test-retest reliability (all *r*s > 0.20) and between-subject consensus (all ICCs > 0.60) [see supplementary text section S2 and Fig. S3]. Eight traits with low factorability (mean absolute correlations with all other traits < 0.3) were excluded from the EFA (see Fig. S4; including these eight traits did not change the results from EFA).

As recommended (*67–69*), we applied parallel analysis to determine the optimal number of factors to retain in the EFA (see supplementary text section S3). Common factor parallel analysis with 5,000 Monte Carlo simulations showed that four factors explained the underlying structure of our dataset (we also obtained estimations with three other methods for comparison such as the optimal coordinates index, see Fig. S5 and supplementary text section S3). EFA was thus applied to extract four factors (using the minimal residual method) and the solutions were rotated with oblimin for interpretability. The four factors each explained 31%, 31%, 11%, and 12% of the common variance in the data (85% in total; 87% in total if five factors were extracted). Figure 2 shows the factor loadings of all traits on the four dimensions. These four dimensions represented attributions regarding warmth [reversed], competence, female-stereotype, and youth-stereotype from faces. Since oblique rotation allowed factors to be correlated with one another, the four dimensions turned out to be weakly correlated (r_{12} = -0.15, r_{13} = -0.33, r_{14} = -0.23, r_{23} = 0.21, r_{24} = 0.33, r_{34} = 0.12).

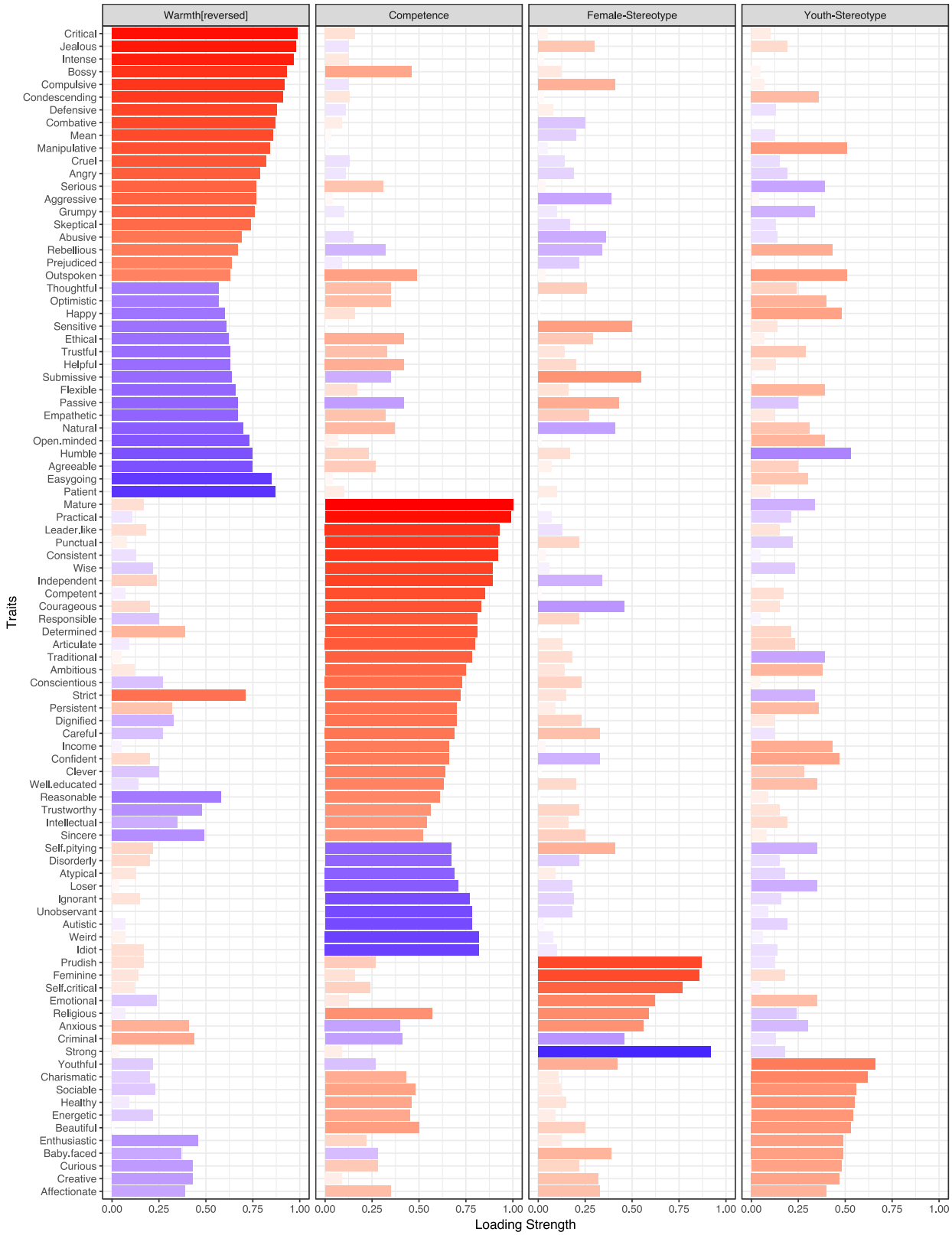

Warmth[reversed]
Competence
Female-Stereotype
Youth-Stereotype
Traits
Critical
Jealous
Intense
Bossy
Compulsive
Condescending
Defensive
Combative
Mean
Manipulative
Cruel
Angry
Serious
Aggressive
Grumpy
Skeptical
Abusive
Rebellious
Prejudiced
Outspoken
Thoughtful
Optimistic
Happy
Sensitive
Ethical
Trustful
Helpful
Submissive
Flexible
Passive
Empathetic
Natural
Open.minded
Humble
Agreeable
Easygoing
Patient
Mature
Practical
Leader.like
Punctual
Consistent
Wise
Independent
Competent
Courageous
Responsible
Determined
Articulate
Traditional
Ambitious
Conscientious
Strict
Persistent
Dignified
Careful
Income
Confident
Clever
Well.educated
Reasonable
Trustworthy
Intellectual
Sincere
Self.pitying
Disorderly
Atypical
Loser
Ignorant
Unobservant
Autistic
Weird
Idiot
Prudish
Feminine
Self.critical
Emotional
Religious
Anxious
Criminal
Strong
Youthful
Charismatic
Sociable
Healthy
Energetic
Beautiful
Enthusiastic
Baby.faced
Curious
Creative
Affectionate
0.00
0.25
0.50
0.75
1.00
Loading Strength

Fig. 2 Comprehensive trait attributions from faces are described by four dimensions.

Each column plots the strength of the factor loadings (x-axis, absolute value) across all 92 traits (y-axis) that were used for the exploratory factor analysis (including all 100 traits revealed the same four dimensions). Color indicates the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

Comparison between the four-dimensional space and existing theories

To understand whether our discovery of a four-dimensional space as distinct from existing theories might be due to methodological differences, we reanalyzed the dimensionality of our data using principal components analysis, as this was the method used in some previous studies (*8*, *9*, *35*). We verified that the same four dimensions emerged when a principal components analysis with varimax rotation was applied: warmth, competence, female-stereotype, and youth-stereotype (Fig. S6) [the first four principal components without rotation accounted for 52%, 21%, 7%, and 5% of the variance in our data, 86% in total; the fifth accounted for 2%].

Next, we investigated the possibility that our four-dimensional space had been missed in the literature because previous studies only examined a very limited set of traits. Here, we inspected two small subsets of our data from Study 1 that consisted of 13 traits corresponding to those examined in (*8*) and (*35*). Indeed, our four-dimensional space was not evident when we restricted the analysis to these two small subsets of traits; instead, we reproduced the two- and three-dimensional spaces found in the two previous studies, respectively (Table S2). In fact, the smallest subset of traits that yield our four-dimensional space was comprised of the 18 traits that showed the highest factor loadings on the four dimensions (Table S3), a minimal set of traits that could be used most efficiently in future studies.

Finally, we directly compared how well our four-dimensional framework and the existing alternative frameworks characterized trait attributions from faces. For each of the three different frameworks, we identified the traits that best captured the meanings of their dimensions in our 100 selected traits (e.g., agreeable, leader-like, feminine, youthful for our framework), whose ratings were then used to predict attributions of all other selected traits (see Fig. S7). Our four-dimensional framework better explained the variance for 72% of the trait attributions than existing theories [mean adjusted R-squared across all predictions was 0.75 for our framework, 0.69 for the framework from (*35*), and 0.63 for the framework from (*8*)].

Reproduction of the four-dimensional space across different countries

Prior studies not only used a considerably restricted set of traits, but also found discrepant dimensions for participants from different cultures (*8*, *9*, *35*, *38*, *51*). To address the generalizability of our findings, we conducted a second large-scale preregistered study in which we collected attributions of 80 traits (a subset of our 100 selected traits) from our 100 faces across seven different countries and regions of the world (see supplementary materials and methods sections M4 and M6). Data across all seven samples showed satisfactory within-subject test-retest reliability and between-subject consensus for conducting subsequent analyses (see supplementary text sections S4 and S5).

We first analyzed the aggregate-level data for each sample. We began by asking whether these seven samples shared a similar correlation structure (the Pearson correlation matrix across trait attributions) with the sample in Study 1. As recommended (*42*, *53*, *70*), we performed representational similarity analysis (RSA) [Fisher z-transformation was applied before computing the Pearson correlation between correlation matrices]. We found high

representational similarity between the sample in Study 1 and all seven samples in Study 2 (*RSA* = 0.96, 95% CI [0.95, 0.96] for North America; *RSA* = 0.92, 95% CI [0.91, 0.92] for Latvia; *RSA* = 0.85, 95% CI [0.84, 0.86] for Peru; *RSA* = 0.85, 95% CI [0.84, 0.86] for the Philippines; *RSA* = 0.75, 95% CI [0.74, 0.77] for India; *RSA* = 0.83, 95% CI [0.82, 0.84] for Kenya; *RSA* = 0.86, 95% CI [0.85, 0.87] for Gaza).

These high RSAs strongly suggest that a similar psychological space underlies face impressions across different cultures (although we emphasize that all our samples were probably westernized to some extent). To understand this psychological space, we again performed parallel analysis to determine the optimal number of dimensions for each sample. We found that a four-dimensional space best described the data in five samples (North America, Latvia, Peru, the Philippines, India) and a three-dimensional space best described the data in the other two samples (Kenya and Gaza) [see Fig. 3A and Fig. S8]. EFA was then applied to extract these dimensions and the solutions were rotated with oblimin for interpretability, as in Study 1. Figure S9 plots the standardized factor loadings across all traits for the seven participant samples. Critically, the four dimensions found in North America, Latvia, Peru, and the Philippines described attributions regarding warmth, competence, female-stereotype, and youth-stereotype, reproducing the finding from Study 1 (Fig. S9A-D). These same four dimensions were also found for Kenya if four factors were extracted (Fig. S9F). To further quantify the similarity between dimensions across samples, we computed Tucker indices of factor congruence (the cosine distance between pairs of factor loadings). Results confirmed that the four-dimensional space was reproduced in the samples from North America, Latvia, Peru, the Philippines, and Kenya when four factors were

extracted (all diagonal indices > 0.4); whereas a subset of three dimensions of our four-dimensional space were reproduced in the other two samples (Fig. 3B).

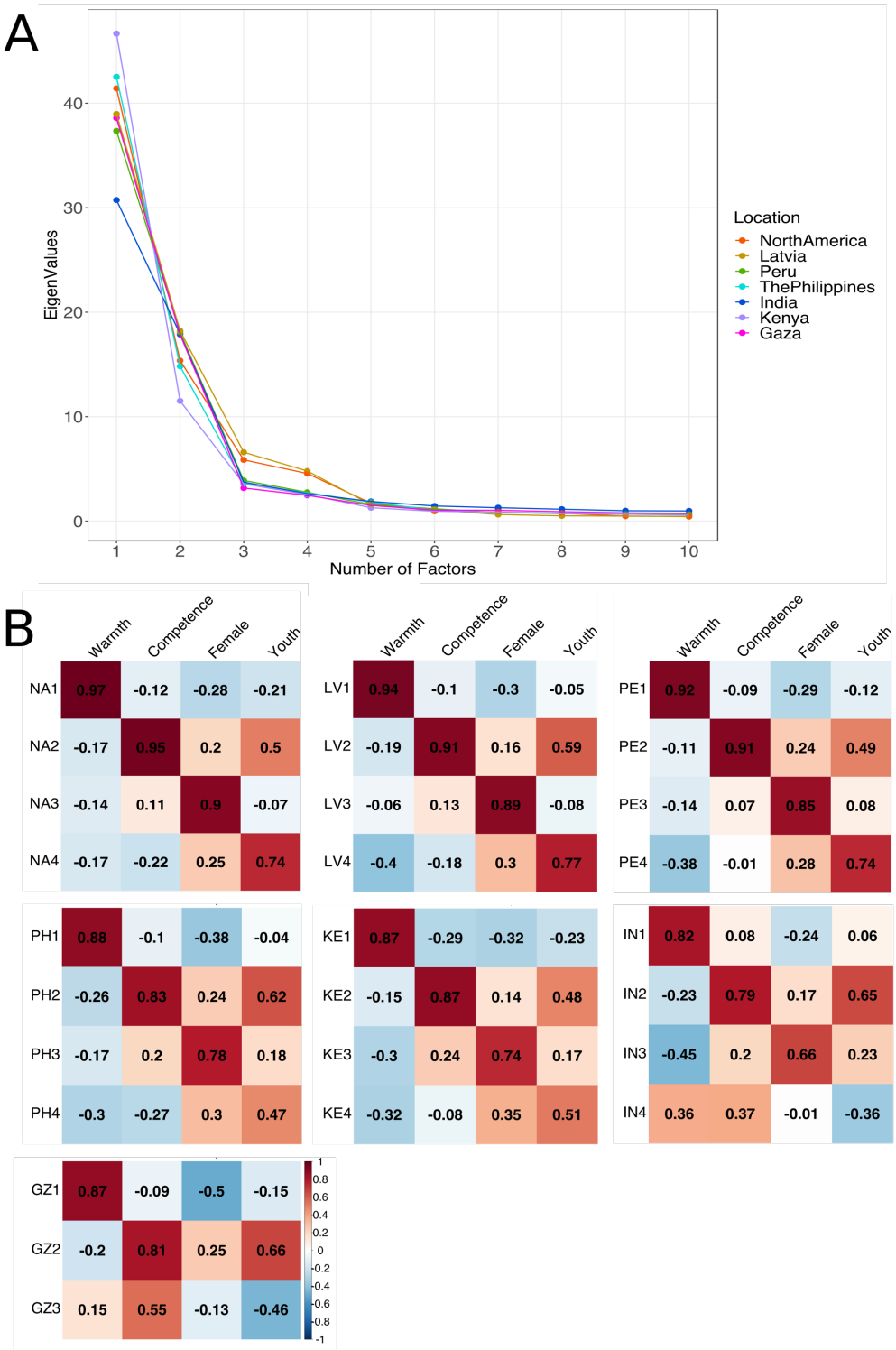


Fig. 3 Dimensionality of trait attributions from faces across samples from different countries.

(**A**) Eigenvalue decomposition. The horizontal axis indicates the first 10 factors (the plotted ordered eigenvalues for the rest of the 80 factors could be approximated with a straight line and were therefore omitted from the graph). The dots plot the eigenvalues of corresponding factors in the samples from North America (red), Latvia (yellow), Peru (green), the Philippines (turquoise), India (blue), Kenya (purple), and Gaza (magenta). (**B**) Tucker indices of factor congruence. Columns indicate the four dimensions found in Study 1. Rows indicate the dimensions found in the samples from North America (NA), Latvia (LV), Peru (PE), the Philippines (PH), Kenya (KN), India (IN), and Gaza (GZ). The numbers report the Tucker indices. The color scale shows the sign and strength of the indices.

Reproduction of the four-dimensional space within individual participants

Although we have reproduced the four-dimensional space in samples from different countries and regions, we have not ruled out the possibility that this space might be an artifact of aggregating data across participants. Could the same four-dimensional space be reproduced in a single participant? This important question has also seldom been addressed in prior work, which has relied on data aggregated across participants to derive the dimensions of a psychological space. To address this question, we turned to the analysis of individual-level data, since we had collected complete datasets in each of the participants in Study 2 (requiring approximately 10 hours of testing per participant spread out across ten days; see supplementary materials and methods section M6). We first performed RSA to investigate whether single participants (n = 86 who had complete datasets after data exclusion, see supplementary text section S4) shared the correlation structure of Study 1. RSAs varied considerably across participants (range = [0.14, 0.85], M = 0.56, SD = 0.16) and, as expected, were attenuated by data quality as assessed by within-subject test-retest reliability (Fig. 4A and B).

We then performed parallel analysis, as preregistered, to determine the dimensionality of trait attributions from each participant. A four-dimensional space was most common across all individual participants (Fig. 4C). Again, the discovery of a four-dimensional space was attenuated by data quality (four-dimensional spaces were found for data with higher test-retest reliability than data that produced other-dimensional spaces [unpaired t-test $t(34.57) = 3.29$, $p = 0.001$]). To inspect whether the four-dimensional spaces discovered in single participants ($n = 22$; cf. Figure 4C) were the same as those derived from aggregated data, we computed the Tucker indices of factor congruence (between the four dimensions discovered in Study 1 and the four dimensions discovered in individual participants in Study 2). Figure S10 summarizes the Tucker indices for all twenty-two participants, among whom eight participants' data largely reproduced the same four-dimensional space as found in aggregated data (the absolute value of all diagonal indices > 0.4).

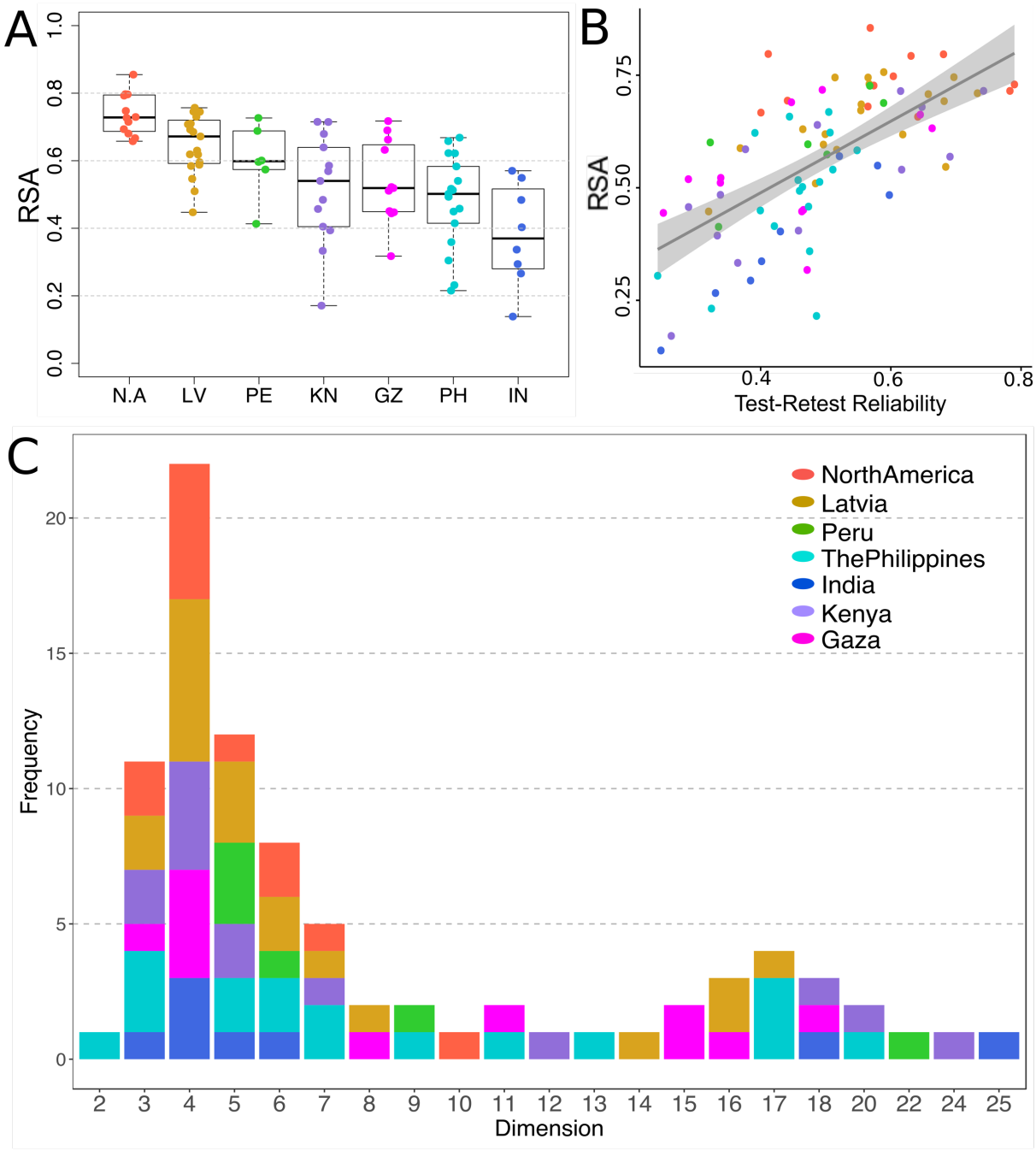


Fig. 4 RSAs and Dimensionality of trait attributions from faces in individual-level data.

(**A**) Boxplot of RSAs. Color indicates different countries and regions (as in Fig. 3A): red for North America (N.A), yellow for Latvia (LV), green for Peru (PE), purple for Kenya (KN), magenta for Gaza (GZ), turquoise for the Philippines (PH), and blue for India (IN). Dots plot the representational similarity between aggregate-level data from

Study 1 and individual-level data from each of the 86 participants in Study 2 who had complete data for attributions of all 80 traits after data exclusion (see supplementary text section S4). Minimum (bottommost line), first quartile (box bottom), median (line in box), third quartile (box top), and maximum (topmost line) of RSAs are presented for each country/region. (**B**) Correlation between test-retest reliability and representational similarity ($R = 0.66$, $p < 0.001$). Color indicates different countries and regions as in (**A**). Each dot plots an individual's test-retest reliability (x-axis) and the individual's representational similarity with the aggregate-level data in Study 1 (y-axis). (**C**) Distribution of the number of dimensions determined by parallel analysis from individual-level data across 86 participants. Color indicates different countries and regions as in (**A**).

Discussion

Humans rapidly and automatically attribute a wide range of traits to others based on their faces (*11*, *54*, *71–74*). These attributions are pervasive and consequential in everyday life (*28–31*, *75–81*). Although we use a large number of different words for these attributions (Fig. S1A) [see also (*8*, *9*)], it has long been thought that the psychological space describing attributions from faces is in fact quite low-dimensional. However, all prior studies have examined only a very limited set of trait attributions, and have produced discrepant findings (*8*, *9*, *35*, *37*, *38*, *40*, *51*), leaving the true nature of the underlying psychological dimensions unclear.

By administering a much more comprehensive set of traits than all prior studies (Fig. 1 and Fig. S1) in two large-scale, pre-registered studies (supplementary materials and methods sections M3-M6), we show that trait attributions from faces are best described by a novel four-dimensional space (Fig. 2). This four-dimensional space was largely reproduced across different countries and regions of the world, even using different languages (Spanish in Peru) [Fig. 3], as well as across many individual participants (although this was more difficult to assess, due to data quality) [Fig. 4 and Fig. S10].

Our discovery of this novel four-dimensional space challenges existing theories and opens a new set of questions. We showed that our divergence from previous findings was not due simply to methodological differences (Fig. S6). Instead, previous studies failed to uncover this four-dimensional space because they investigated a set of traits that is too limited (Table S2). Our four-dimensional space better characterized the variance across trait attributions than do any of the existing theories (Fig. S7), and makes specific recommendations for the trait and face stimuli that could be used in future studies—since it is practically very challenging to administer our complete set of 100 traits, a subset could be selected according to how well they represent the four-dimensional space (Table S3).

Our findings dovetail with a large literature on the dimensionality of social cognition. This literature theorizes that warmth and competence are two universal dimensions of social cognition, which arose in the evolution of social behavior (*82*). For example, when encountering a stranger, an individual needs to determine, first, the intentions of the stranger (warmth), and then the stranger's ability to execute those intentions (competence). In our study, warmth and competence were the first two dimensions of trait attributions from faces (Fig. 2), integrating frameworks describing face impressions with those in the social cognition literature.

While the four-dimensional space we found was largely reproduced across countries, we also noticed some variation across samples and individuals (Fig. 3 and Fig. 4), raising the possibility that the psychological space for face impressions might be modified by culture and individual differences. However, we refrain from drawing any strong conclusions about cultural differences in our research. It is notoriously difficult to assure specific cultural exposure for participants, and we make no such claims here. Instead, our Study 2 was intended to extend the generalizability of

our findings by providing a more culturally diverse participant set, and to collect dense individual-level data. The differences that we found in dimensional structure in some countries should be considered preliminary results that could motivate larger-scale studies focused on cultural differences in the future. Similarly, the exploration of individual differences (and their possible correlations with demographic, state, or trait variables in that individual) will require future studies that collect much denser, and longitudinal data.

Our study has an important limitation in that we only included faces that were white, frontal, with direct gaze, with neutral expressions, and without any glasses or hats obscuring the face. These criteria made it possible to sample stimuli that maximized the variability in facial structure in a homogenous manner (Fig. 1). However, it also precluded the investigation of a host of contextual effects that operate in everyday life. Since a large literature has documented such effects [e.g., of race, or facial expression] (*15*, *83*, *84*), a clear future direction will be to understand how the four dimensions we discovered might be influenced by such additional effects. An important question for future research is whether more diverse face stimuli, as well as faces in ambient photos, would only reveal modulations of the four-dimensional space we discovered here, or would uncover additional or new dimensions altogether.

A proximal explanation for the present findings must ultimately reside in the neural mechanisms that produce the attributions people make (*85*). Future studies using neuroimaging should investigate whether attributions of different categories of traits engage different neural networks, and whether the neural encoding of face impressions is organized in the same four dimensions we discovered here.

References and Notes:

1. O. Wiles, A. S. Koepke, A. Zisserman, in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss, Eds. (Springer International Publishing, Cham, 2018; http://link.springer.com/10.1007/978-3-030-01261-8_41), vol. 11217, pp. 690–706.

2. M. A. Stefanone, Z. Yue, Z. Toh, A social cognitive approach to traditional media content and social media use: Selfie-related behavior as competitive strategy. *New Media & Society*. **21**, 317–335 (2019).

3. J. Chae, Virtual makeover: Selfie-taking and social media use increase selfie-editing frequency through social comparison. *Computers in Human Behavior*. **66**, 370–376 (2017).

4. A. J. O'Toole, C. D. Castillo, C. J. Parde, M. Q. Hill, R. Chellappa, Face Space Representations in Deep Convolutional Neural Networks. *Trends in Cognitive Sciences*. **22**, 794–809 (2018).

5. C. J. Parde, Y. Hu, C. Castillo, S. Sankaranarayanan, A. J. O'Toole, Social Trait Information in Deep Convolutional Neural Networks Trained for Face Identification. *Cognitive Science*. **43** (2019), doi:doi.org/10.1111/cogs.12729.

6. M. Mermillod, P. Bonin, L. Mondillon, D. Alleysson, N. Vermeulen, Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*. **73**, 2522–2531 (2010).

7. F. Xu, J. Zhang, J. Z. Wang, Microexpression Identification and Categorization Using a Facial Dynamics Map. *IEEE Transactions on Affective Computing*. **8**, 254–267 (2017).

8. N. N. Oosterhof, A. Todorov, The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*. **105**, 11087–11092 (2008).

9. C. A. M. Sutherland, X. Liu, L. Zhang, Y. Chu, J. A. Oldmeadow, A. W. Young, Facial First Impressions Across Culture: Data-Driven Modeling of Chinese and British Perceivers' Unconstrained Facial Impressions. *Pers Soc Psychol Bull*. **44**, 521–537 (2018).

10. J. Na, S. Kitayama, Spontaneous Trait Inference Is Culture-Specific: Behavioral and Neural Evidence. *Psychological Science*. **22**, 1025–1032 (2011).

11. A. D. Engell, J. V. Haxby, A. Todorov, Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience*. **19**, 1508–1519 (2007).

12. E. J. Cogsdill, A. T. Todorov, E. S. Spelke, M. R. Banaji, Inferring Character From Faces: A Developmental Study. *Psychol Sci*. **25**, 1132–1139 (2014).

13. R. B. Adams Jr, A. J. Nelson, J. A. Soto, U. Hess, R. E. Kleck, Emotion in the neutral face: A mechanism for impression formation? *Cognition and Emotion*. **26**, 431–441 (2012).

14. N. O. Rule, N. Ambady, Democrats and Republicans Can Be Differentiated from Their Faces. *PLOS ONE*. **5**, e8733 (2010).

15. A. Todorov, *Face value: The irresistible influence of first impressions* (Princeton University Press, 2017).

16. S. Porter, L. England, M. Juodis, L. ten Brinke, K. Wilson, Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*. **40**, 171–177 (2008).

17. I. S. Penton-Voak, N. Pound, A. C. Little, D. I. Perrett, Personality Judgments from Natural and Composite Facial Images: More Evidence For A "Kernel Of Truth" In Social Perception. *Social Cognition*. **24**, 607–640 (2006).

18. N. O. Rule, J. V. Garrett, N. Ambady, On the Perception of Religious Group Membership from Faces. *PLOS ONE*. **5**, e14241 (2010).

19. N. O. Rule, J. M. Moran, J. B. Freeman, S. Whitfield-Gabrieli, J. D. E. Gabrieli, N. Ambady, Face value: Amygdala response reflects the validity of first impressions. *NeuroImage*. **54**, 734–741 (2011).

20. C. Y. Olivola, A. Todorov, Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*. **46**, 315–324 (2010).

21. C. Y. Olivola, F. Funk, A. Todorov, Social attributions from faces bias human choices. *Trends in Cognitive Sciences*. **18**, 566–570 (2014).

22. J.-F. Bonnefon, A. Hopfensitz, W. De Neys, Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences*. **19**, 421–422 (2015).

23. L. Ewing, F. Caulfield, A. Read, G. Rhodes, Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science*. **18**, 327–334 (2015).

24. S. Porter, L. ten Brinke, C. Gustaw, Dangerous decisions: the impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*. **16**, 477–491 (2010).

25. J. P. Wilson, N. O. Rule, Hypothetical Sentencing Decisions Are Associated With Actual Capital Punishment Outcomes: The Role of Facial Trustworthiness. *Social Psychological and Personality Science*. **7**, 331–338 (2016).

26. C. Lin, R. Adolphs, R. M. Alvarez, Cultural effects on the association between election outcomes and face-based trait inferences. *PLOS ONE*. **12**, e0180837 (2017).

27. C. Lin, R. Adolphs, R. M. Alvarez, Inferring Whether Officials Are Corruptible From Looking At Their Faces. *Psychological Science*. **29**, 1807–1823 (2018).

28. A. Todorov, Inferences of Competence from Faces Predict Election Outcomes. *Science*. **308**, 1623–1626 (2005).

29. I. V. Blair, C. M. Judd, K. M. Chapleau, The Influence of Afrocentric Facial Features in Criminal Sentencing. *Psychol Sci*. **15**, 674–679 (2004).

30. J. P. Wilson, N. O. Rule, Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*. **26**, 1325–1331 (2015).

31. J. Antonakis, D. L. Eubanks, Looking Leadership in the Face. *Current Directions in Psychological Science*. **26**, 270–275 (2017).

32. A. Todorov, C. Y. Olivola, R. Dotsch, P. Mende-Siedlecki, Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*. **66**, 519–545 (2015).

33. E. Hehman, C. A. M. Sutherland, J. K. Flake, M. L. Slepian, The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*. **113**, 513–529 (2017).

34. L. A. Zebrowitz, First Impressions From Faces. *Current Directions in Psychological Science*. **26**, 237–242 (2017).

35. C. A. M. Sutherland, J. A. Oldmeadow, I. M. Santos, J. Towler, D. Michael Burt, A. W. Young, Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*. **127**, 105–118 (2013).

36. M. Mileva, A. W. Young, R. S. S. Kramer, A. M. Burton, Understanding facial impressions between and within identities. *Cognition*. **190**, 184–198 (2019).

37. J. R. Collova, C. A. M. Sutherland, G. Rhodes, *Journal of Personality and Social Psychology*, in press, doi:10.1037/pspa0000167.

38. H. Wang, C. Han, A. C. Hahn, V. Fasolt, D. K. Morrison, I. J. Holzleitner, L. M. DeBruine, B. C. Jones, A data-driven study of Chinese participants' social judgments of Chinese faces. *PLOS ONE*. **14**, e0210315 (2019).

39. D. Oh, R. Dotsch, J. Porter, A. Todorov, Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General* (2019).

40. J. K. South Palomares, C. A. M. Sutherland, A. W. Young, Facial first impressions and partner preference models: Comparable or distinct underlying structures? *Br J Psychol*. **109**, 538–563 (2018).

41. S. Y. Xie, J. K. Flake, E. Hehman, Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*. **117**, 364–385 (2019).

42. R. M. Stolier, E. Hehman, M. D. Keller, M. Walker, J. B. Freeman, The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*. **115**, 9210–9215 (2018).

43. C. A. M. Sutherland, L. E. Rowley, U. T. Amoaku, E. Daguzan, K. A. Kidd-Rossiter, U. Maceviciute, A. W. Young, Personality judgments from everyday images of faces. *Frontiers in Psychology*. **6** (2015), doi:10.3389/fpsyg.2015.01616.

44. C. A. M. Sutherland, A. W. Young, C. A. Mootz, J. A. Oldmeadow, Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *Br J Psychol*. **106**, 186–208 (2015).

45. M. Oliveira, T. Garcia-Marques, R. Dotsch, L. Garcia-Marques, Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*. **49**, 888–902 (2019).

46. B. C. Jones, L. M. DeBruine, J. K. Flake, B. Aczel, M. Adamkovic, R. Alaei, S. Alper, et al., To Which World Regions Does the Valence-Dominance Model of Social Perception Apply?. *PsyArXiv* (2018), doi:10.31234/osf.io/n26dy.

47. C. A. M. Sutherland, A. W. Young, G. Rhodes, Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *Br J Psychol*. **108**, 397–415 (2017).

48. C. A. M. Sutherland, J. A. Oldmeadow, A. W. Young, Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*. **157**, 257–267 (2016).

49. R. J. W. Vernon, C. A. M. Sutherland, A. W. Young, T. Hartley, Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*. **111**, E3353–E3361 (2014).

50. L. H. Stewart, S. Ajina, S. Getov, B. Bahrami, Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General* (2012), , doi:10.1037/a0027950.

51. E. Hehman, R. M. Stolier, J. B. Freeman, J. K. Flake, S. Y. Xie, Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Soc Personal Psychol Compass*. **13**, e12431 (2019).

52. G. Saucier, L. R. Goldberg, Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality*. **10**, 61–77 (1996).

53. R. M. Stolier, E. Hehman, J. B. Freeman, Conceptual structure shapes a common trait space across social cognition. *PsyArXiv* (2019), doi:10.31234/osf.io/5na8m.

54. L. A. Zebrowitz, J. M. Montepare, Social Psychological Face Perception: Why Appearance Matters. *Soc Personal Psychol Compass*. **2**, 1497 (2008).

55. C. P. Said, N. Sebe, A. Todorov, Structural Resemblance to Emotional Expressions Predicts Evaluation of Emotionally Neutral Faces (2009).

56. N. O. Rule, N. Ambady, K. C. Hallett, Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*. **45**, 1245–1251 (2009).

57. C. Y. Olivola, A. Todorov, Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *J Nonverbal Behav*. **34**, 83–110 (2010).

58. A. Todorov, P. Mende-Siedlecki, R. Dotsch, Social judgments from faces. *Current Opinion in Neurobiology*. **23**, 373–380 (2013).

59. P. F. Secord, W. F. Dukes, W. Bevan, Personalities in faces: I. An experiment in social perceiving. *Genetic Psychology Monographs*. **49**, 231–270 (1954).

60. G. W. Allport, H. S. Odbert, *Psychological monographs*, in press.

61. P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 135–146 (2017).

62. L. DeBruine, B. Jones, Face Research Lab London Set (2017), , doi:10.6084/m9.figshare.5047666.v3.

63. D. S. Ma, J. Correll, B. Wittenbrink, The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*. **47**, 1122–1135 (2015).

64. O. Chelnokova, B. Laeng, M. Eikemo, J. Riegels, G. Løseth, H. Maurud, F. Willoch, S. Leknes, Rewards of beauty: the opioid system mediates social motivation in humans. *Molecular Psychiatry*. **19**, 746–747 (2014).

65. D. E. King, Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, 41755–1758 (2009).

66. Michael Quinn Patton, *Qualitative Research & Evaluation Methods* (SAGE Publications, Thousand Oaks, CA, ed. 3rd, 2002; https://us.sagepub.com/en-us/nam/qualitative-research-evaluation-methods/book232962).

67. W. R. Zwick, W. F. Velicer, Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*. **99**, 432–442 (1986).

68. Ö. Çokluk, D. Koçak, Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors. *Educational Sciences: Theory & Practice*. **16**, 537–551 (2016).

69. R. Pearson, D. Mundfrom, A. Piccone, A Comparison of Ten Methods for Determining the Number of Factors in Exploratory Factor Analysis. **39**, 15 (2013).

70. J. A. Brooks, J. B. Freeman, Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*. **2**, 581–591 (2018).

71. K. L. Ritchie, R. Palermo, G. Rhodes, Forming impressions of facial attractiveness is mandatory. *Sci Rep*. **7**, 1–8 (2017).

72. R. Hassin, Y. Trope, Facing Faces: Studies on the Cognitive Aspects of Physiognomy, 16 (2000).

73. K. Dobs, L. Isik, D. Pantazis, N. Kanwisher, How face perception unfolds over time. *Nature Communications*. **10**, 1258 (2019).

74. M. Bar, M. Neta, H. E. Linz, Very first impressions. *Emotion*. **6**, 269–278 (2006).

75. C. Rezlescu, B. Duchaine, C. Y. Olivola, N. Chater, Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior. *PLOS ONE*. **7**, e34293 (2012).

76. M. van 't Wout, A. G. Sanfey, Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*. **108**, 796–803 (2008).

77. K. A. Valentine, N. P. Li, L. Penke, D. I. Perrett, Judging a Man by the Width of His Face: The Role of Facial Ratios and Dominance in Mate Choice at Speed-Dating Events. *Psychological Science*. **25**, 806–811 (2014).

78. A. Genevsky, B. Knutson, Neural Affective Mechanisms Predict Market-Level Microlending. *Psychol Sci*. **26**, 1411–1422 (2015).

79. A. I. Gheorghiu, M. J. Callan, W. J. Skylark, Facial appearance affects science communication. *Proc Natl Acad Sci USA*. **114**, 5970–5975 (2017).

80. D. S. Berry, L. Zebrowitz-McArthur, What's in a Face?: Facial Maturity and the Attribution of Legal Responsibility. *Pers Soc Psychol Bull*. **14**, 23–33 (1988).

81. L. A. Zebrowitz, S. M. McDonald, The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law Hum Behav*. **15**, 603–623 (1991).

82. S. T. Fiske, A. J. C. Cuddy, P. Glick, Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*. **11**, 77–83 (2007).

83. D. T. Levin, M. R. Banaji, Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*. **135**, 501–512 (2006).

84. L. A. Zebrowitz, M. Kikuchi, J.-M. Fellous, Facial Resemblance to Emotions: Group Differences, Impression Effects, and Race Stereotypes. *J Pers Soc Psychol*. **98**, 175–189 (2010).

85. J. A. Brooks, J. B. Freeman, Neuroimaging of person perception: A social-visual interface. *Neuroscience Letters*. **693**, 40–43 (2019).

86. E. Hehman, S. Y. Xie, E. K. Ofosu, G. Nespoli, Assessing the point at which averages are stable: A tool illustrated in the context of person perception (2018), doi:10.31234/osf.io/2n6jq.

Acknowledgments: We thank Dean Mobbs, R. Michael Alvarez, Antonio Rangel, Clare Sutherland, Uri Maoz, and William Revelle for their valuable input, Remya Nair and Christopher J. Birtja for technology support, and Becky Santora for helping with testing participants in foreign locations through Digital Divide Data. **Funding:** Funded in part by NSF grants BCS-1840756 and BCS-1845958, and the Carver Mead New Adventures Fund.

Author contributions: C.L. and R.A. developed the study concept and designed the study; C.L. and U.K. prepared experimental materials; R.A. supervised the experiments and analyses; C.L. performed and supervised data collection; C.L. and U.K. performed data analyses; C.L. and R.A.

drafted the manuscript; all authors revised and reviewed the manuscript and approved the final manuscript for submission.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data, codes, and materials are available at Open Science Framework: http://bit.ly/osfface1 and http://bit.ly/osfface2.

Supplementary Materials

Materials and Methods

M1. Trait Stimuli

Our goal was to sample the most comprehensive list of trait-words that are used to describe people based on their faces. We derived a final set of 100 traits through a series of combinations and filters. These 100 traits were further verified to be representative of words that people freely generate for our face stimuli (see section M7).

To derive the final trait set, we first gathered an inclusive list of 482 adjectives and 6 nouns that describe a person's demographic characteristics, physical appearance, social evaluative qualities, personality, and emotional traits, from multiple sources (*8, 9, 15, 26, 28, 32, 33, 52–60*).

Many of the 482 adjectives had similar or opposite meanings. To avoid redundancy while conserving semantic variability, we sampled these adjectives according to three criteria: their semantic similarity, clarity in meaning, and frequency in usage. For those with similar meanings, clarity was the second selection criterion (the one with the highest clarity was retained). For those with similar meanings and the same clarity, usage frequency was the third selection criterion (the one with the highest usage frequency was retained).

To quantify the semantic similarity between these 482 adjectives, we represented each of them with a vector of 300 computationally extracted semantic features (describing word embeddings and text classification) using a state-of-the-art neural network provided within the FastText library (*61*) that had been trained on Common Crawl data of 600 billion words to predict the identity of a word given a context. We then applied hierarchical agglomerative clustering (HAC) on the word vectors based on their cosine distances to visualize their semantic similarities.

To quantify clarity of meaning, we obtained ratings of clarity from an independent set of participants tested via MTurk (N = 31, 17 males, Age (M = 36, SD = 10)).

To quantify usage frequency, we obtained the average monthly Google search frequency for the bigram of each adjective (i.e., the adjective together with the word "person" added after it) using the keyword research tool Keywords Everywhere (https://keywordseverywhere.com/).

Based on the three criteria, the 482 adjectives were reduced to 94 adjectives. These 94 adjectives together with the 6 nouns of additional demographic and health characteristics (education, income, sexuality, autism) and frequently used derogatory words (idiot, loser) comprised our final set of 100 traits. The preregistration of trait stimuli can be accessed at Open Science Framework (http://bit.ly/osfpre1).

M2. Face Stimuli

Our goal was to derive a representative set of face images that are of excellent quality (e.g., clear, frontal) and diverse facial structures (maximizing variability in facial structure while controlling for factors such as race, expressions, angles, gaze, and background, which our present project does not intend to investigate).

We first combined 909 high-resolution photographs of male and female faces from three publicly available face databases: the Oslo Face Database (*64*), the Chicago Face Database (*63*), and the Face Research Lab London Set (*62*). We then excluded faces that were not front-facing, not with direct-gaze, and with glasses or other adornments obscuring the face. We then further

restricted ourselves to photographs of Caucasian adults and neutral expression, because we were not interested in investigating race or emotion variables in this study. This yielded a set of 426 faces from the three databases.

To further reduce the size of the stimulus set while conserving variability in facial structure, we sampled from the 426 faces using maximum variation sampling. For each image, the face region was first detected and cropped using the dlib library (*65*), and then represented with a vector of 128 computationally extracted facial features (for face recognition), using a state-of-the-art neural network provided within the dlib library that had been trained to identify individuals across millions of faces (of all different aspects and races) with very high accuracy (*65*). Next, we sampled 50 female faces and 50 male faces that respectively maximized the sum of the Euclidean distances between their face vectors. Specifically, a face image was first randomly selected from the female or male sampling set, and then other images of the same gender were selected so that each new selected image had the farthest Euclidean distance from the previously selected images. We repeated this procedure with 10,000 different initializations and selected the sample with the maximum sum of Euclidean distances. We repeated the whole sampling procedure 50 times to ensure convergence of the final sample.

All 100 final faces were clear, frontal, with neutral expression, and presented at the center of the images with the eyes at the same height across images. All photos included the face, neck, and hair, were colored, had a standard grey background, and were cropped to a standard size and shape. The preregistration of face stimuli and the final list of faces can be accessed at Open Science Framework (http://bit.ly/osfface1).

M3. Participants (Study 1)

The study was approved by the Institutional Review Board of the California Institute of Technology and informed consent was obtained from all participants. We predetermined our sample size to be 60 participants per trait based on a recent study that investigated the point of stability for trait attributions from faces (*86*). That study (*86*) analyzed a dataset containing 698,829 ratings from 6,593 participants for 3,353 facial stimuli and 24 traits (*33*), and found that a stable average rating could be obtained in a sample of 18 to 42 participants across the 24 traits (ratings were elicited on a 7-point Likert scale, the acceptable corridor of stability was +/- 0.5, and the confidence level was 95%). Based on these findings, we preregistered our sample size to be 60 participants for each trait (see preregistration at http://bit.ly/osfpre1).

Participants were recruited via Amazon Mechanical Turk (N = 1,500 (800 males), Age(M = 38 years, SD = 11), the median of educational attainment was "some post-high-school, no bachelor's degree"). All participants were required to be white, native English speakers, located in the U.S., 18 years old or older, with normal or corrected-to-normal vision, with an educational attainment of high school or above, and with a good MTurk participation history (approval rating $\geq$ 95%).

We also collected data about whether our participants were currently being treated for psychiatric or neurological illness. The majority of our participants (79.7%) were not currently being treated for any psychiatric or neurological illness. The rest were currently being treated for depression (9.8%), bipolar disorder (1.3%), anxiety or panic disorder (11.2%), obsessive compulsive disorder (0.9%), post-traumatic stress disorder (1.3%), autism spectrum disorder (0.3%), learning disability (0.1%), attention deficit disorder (0.9%), alcohol or drug addiction (1.0%), personality disorder (0.5%), dissociative disorder (0.1%), epilepsy (0.2%), and brain injury

(0.1%). All dimensional analyses reported in the main text were repeated on those 79.7% participants who were not currently being treated for any psychiatric or neurological illness, and the results corroborated findings from the full dataset: the same eight traits were found to have low factorability and therefore removed from subsequent dimensional analyses; parallel analysis together with Cattell's scree test and optimal coordinates index indicated that the optimal number of factors was four; these four factors extracted with EFA were identical to the four dimensions reported in Study 1 (Tucker indices of factor congruence = 1.00, 1.00, 0.99, 0.99).

M4. Participants (Study 2)

The study was approved by the Institutional Review Board of California Institute of Technology and informed consent was obtained from all participants. We preregistered to recruit participants through Digital Divide Data, a social enterprise that delivers research services, in seven countries/regions of the world: North America (U.S. and Canada), Latvia, Peru, the Philippines, India, Kenya, and Gaza. All participants were required to be between 18-40 years old, proficient in English (except participants in Peru), have been educated and completed at minimum high school, have been trained in basic computer skills, and have never visited or lived in western-culture countries (except participants in North America and Latvia). In addition, we aimed to have a roughly equal sex ratio of participants in all locations.

The sample size for each location was predetermined to be 30 participants. This sample size was determined based on two criteria: first, the sample size should be large enough to ensure stable average trait ratings [for a corridor of stability of +/- 1.00 and a level of confidence of 95%, the point of stability ranged from 5 to 11 participants across 24 traits (*86*)]; second, the sample size should be feasible to accrue at all seven locations given the requirements mentioned above and the availability of participants for paying multiple visits to complete all our experiment sessions over a 10-day period (see preregistration at http://bit.ly/osfpre2).

As planned, 30 individuals (15 females and 15 males) in each of the seven locations participated in our study (Age (M = 26, SD = 4) for North America; Age (M = 28, SD = 5) for Latvia; Age (M = 22, SD = 3) for Peru; Age (M = 25, SD = 4) for the Philippines; Age (M = 27, SD = 6) for India; Age (M = 24, SD = 2) for Kenya; and Age (M = 26, SD = 5) for Gaza). All participants were confirmed to meet the requirements mentioned above.

M5. Procedures (Study 1)

All experiments were completed online via MTurk. Considering the large amount of time it would take for a participant to complete ratings for all one hundred traits and one hundred faces, we divided the experiment into 25 modules (the 100 traits were randomly shuffled once and divided into 25 modules, each consisting of 4 traits). Each participant completed one module.

To encourage participants to use the full range of the rating scale, all one hundred faces were shown briefly (in five sets of arrays) at the beginning of a module, so that participants had a sense of the range of the faces they were going to rate. In each module, participants rated all faces on each of the four traits (in random order) in the first four blocks, and then in the last (fifth) block they rerated all faces on the trait they were assigned in the first block again, thus providing sparse test-retest data for our traits.

At the beginning of each block, participants were instructed on the trait they were asked to evaluate and were provided with a clear one-sentence definition of the trait (Table S1). Participants viewed the faces one by one in random order and rated each face on a trait using a 7-point Likert

scale. Each face appeared for one second. Participants could enter their ratings as soon as the photo appeared or within four seconds after the photo disappeared. Participants entered their ratings by pressing the number keys on the computer keyboard. The orientation of the Likert scale in each block was randomized across participants. At the end of the experiment, participants completed a brief questionnaire on demographic information (see preregistration at http://bit.ly/osfpre1).

M6. Procedures (Study 2)

All experiments were completed onsite in the Digital Divide Data local offices. Participants in North America, Latvia, the Philippines, India, Kenya, and Gaza completed all experiments in English. Participants in Peru completed all experiments in Spanish. An exact translation of the experiment instructions, trait words, and definitions of the traits from English to Spanish was provided by the Peru office of Digital Divide Data.

Eighty of the 100 traits were used in Study 2—twenty traits were excluded for their low correlations with other traits as found in Study 1 (sarcastic, white, thrifty, shallow, homosexual, nosey, conservative, and reserved), their ambiguity or similarity in meaning as found in feedback from Study 1 (trustful, natural, passive, reasonable, strict, enthusiastic, affectionate, and sincere), and their potential inappropriateness in some cultures (idiot, loser, criminal, and abusive).

Participants in all seven countries/regions followed the same experimental procedures. Each participant provided evaluations on all traits for all faces, of which 20 traits were rated twice for test-retest reliability. The 80 traits were divided into 20 modules, each consisting of 4 distinct traits (the 20 retested traits were first assigned to distinct modules and then the other traits were randomly assigned across modules with the constraint that traits in the same module should be balanced in valence). All participants completed all 20 modules during multiple visits to the local offices in ten business days. Each module consisted of 5 blocks, with the retested trait always shown in the first and last blocks and the other traits shown in random order. The experimental procedure within each module was identical to Study 1.

Both the English and Spanish versions of the experiment instructions, the lists of traits and retested traits, and the definitions of the traits can be accessed at our preregistration (http://bit.ly/osfpre2).

M7. Procedures (Freely generated traits)

The study was approved by the Institutional Review Board of California Institute of Technology and informed consent was obtained from all participants. As preregistered, 30 participants were recruited via MTurk (see preregistration at http://bit.ly/osfpre4); different from the preregistration, we decided to not only include Caucasian participants but included participants of any race (27 participants were white, 3 participants were black).

Participants viewed the 100 faces one by one, each for 1 second. After the disappearance of each face, participants were asked to type in the words (preferably single-word adjectives) that came to mind about the person they've just seen. Participants could type in as many as ten words and were encouraged to type in at least four words (the number of words entered per trial—words entered by a participant for one face—ranged from 0 word [for 8 trials] to 10 words [for 190 trials] with mean = 5 words). There was no time limit for word entering; participants clicked "confirm" to move on to the next trial when they finished entering all the words they wanted to enter for the current trial.

Supplementary Text

S1. Data Processing (Study 1)

Data were excluded following three preregistered criteria: a. Trial-wise deletion if a response was missing or timed out, or if RT was less than 100ms; b. Participant-wise deletion if a participant had more than 10% invalid trials in any block, defined as per (a); c. Block-wise (trait-wise per participant) deletion if all trials in a given block had the same rating (see http://bit.ly/osfpre1).

Following our preregistered data exclusion criteria, of the full sample with a registered size of N = 1,500 participants and L = 750,000 ratings, n = 48 participants and l = 27,491 ratings were excluded from further analysis.

S2. Measures of Within-subject Test-retest Reliability and Between-subject Consensus (Study 1)

Each of the one hundred traits was rated twice for all faces by nonoverlapping subsets of participants (ca. n = 15 per trait). Following our preregistered data analysis plan, we applied linear mixed-effect modeling to assess within-subject test-retest reliability, which adjusted for non-independence in repeated individual ratings by incorporating both fixed effects (that were constant across participants) and random effects (that varied across participants). Ratings from every participant for every face collected at the second time were regressed on those collected at the first time (ca. l = 1,445 pairs of ratings per trait after data exclusion) while controlling for the random effect of participants. As hypothesized in our preregistration, we found that ratings of traits about physical appearance, such as white (r = 0.81), feminine (r = 0.80), strong (r = 0.68), youthful (r = 0.67), baby-faced (r = 0.67), beautiful (r = 0.67) had high within-subject test-retest reliabilities. To our surprise, ratings of autistic also showed a high test-retest reliability (r = 0.64). Ratings of whether the person had low or high income showed the lowest test-retest reliability (r = 0.22).

Following our preregistered data analysis plan, we assessed the between-subject consensus for each trait with intraclass correlation coefficients (ICC(2,k)), using ratings of every face by every participant (ca. n = 58 participants and l = 5,780 ratings per trait after data exclusion). A high intraclass correlation coefficient indicates that the total variance in the ratings is mainly explained by the variance across faces instead of participants. We observed excellent between-subject consensus (ICCs greater than 0.75) for ninety-three of the one hundred traits. Traits with the highest between-subject consensus were those concerning physical appearance, such as feminine, white, youthful, strong, beautiful, and baby-faced. The seven traits with the lowest consensus (self-critical, sarcastic, reserved, anxious, thrifty, shallow, and compulsive) still had good ICCs (ICCs ranged from 0.60 to 0.75) [Fig. S3].

S3. Determination of the Optimal Number of Factors to Retain in EFA

As recommended (*67–69*), we applied parallel analysis to determine the optimal number of factors to retain in EFA (Fig. S5 and Fig. S8). Parallel analysis retains factors that are not simply due to chance by comparing the eigenvalues of the observed data matrix with those of multiple randomly generated data matrices that match the sample size of the observed data matrix. This produces accurate estimations consistently across different conditions (e.g., the distribution properties of the data) (*67–69*).

For completeness, we also obtained estimations based on Kaiser's rule, Cattell's scree test, and the optimal coordinates index, approaches that are commonly used to determine the number

of factors to retain in EFA but that are generally less accurate than parallel analysis (*67–69*). Kaiser's rule retains factors with eigenvalues that are greater than one. Cattell's scree test retains factors to the left of the point from which the plotted ordered eigenvalues could be approximated with a straight line (i.e., retains factors "above the elbow"). The optimal coordinates index provides a non-graphical solution to Cattell's scree test based on linear extrapolation.

S4. Data Processing (Study 2)

Two exclusion criteria were planned in the initial preregistration: a. Trial-wise deletion if a response was missing or timed out, or if RT was less than 100ms; b. Block-wise (trait-wise per participant) deletion if all trials in a given block had the same rating.

To ensure high quality and complete data from individuals, we further registered four exclusion criteria while data collection was underway and data had not yet been analyzed. A. Trial-wise deletion if a rating was missing or timed out, or if RT was less than 400ms; B. Block-wise (trait-wise per participant) deletion if (B1) a block had more than 10% ratings that were missing or with RTs less than 100ms, or (B2) a block had more than 20% ratings with RTs less than 400ms, or (B3) a block had the same rating for all faces; C. Participant-wise deletion if a participant's test-retest reliabilities for more than 25% of the retested traits were more than three standard deviations below the mean test-retest reliability as found in Study 1; D. Participant-wise deletion if a participant's test-retest reliabilities for more than 50% of the retested traits were below 0.20 (see http://bit.ly/osfpre2 and http://bit.ly/osfpre3).

Following criteria A to C, of the full sample with a preregistered size of $N = 30$ participants and $L = 300{,}000$ ratings at each of 7 locations ($N = 210$ total), we excluded from further analysis $n = 1$ participant in India and $l = 24{,}236$ ratings in North America, $l = 2{,}507$ ratings in Latvia, $l = 16{,}366$ ratings in Peru, $l = 3{,}178$ ratings in the Philippines, $l = 14{,}389$ ratings in India, $l = 9{,}117$ ratings in Kenya, and $l = 4{,}096$ ratings in Gaza. Analyses of within-subject test-retest reliability and between-subject consensus were performed using data that were processed with exclusion criteria A to C; criterion D was not applied for those analyses because it imposed a strict lower bound on the within-subject test-retest reliability to ensure data quality, which might lead to an overestimation of the reliability of the data.

Following criteria A to D, thirty-one participants across seven locations were excluded for further analysis ($n = 3$ for North America, $n = 2$ for Latvia, $n = 7$ for Peru , $n = 3$ for the Philippines, $n = 10$ for India, $n = 2$ for Kenya, and $n = 4$ for Gaza). Analyses of representational similarity and dimensionality were performed using data that were processed with exclusion criteria A to D (Fig. 3 and Fig. S8-9). Among those remaining participants, $n = 86$ participants had complete data for the attributions of all 80 traits—data from these 86 participants were used in the individual-level analyses (Fig. 4).

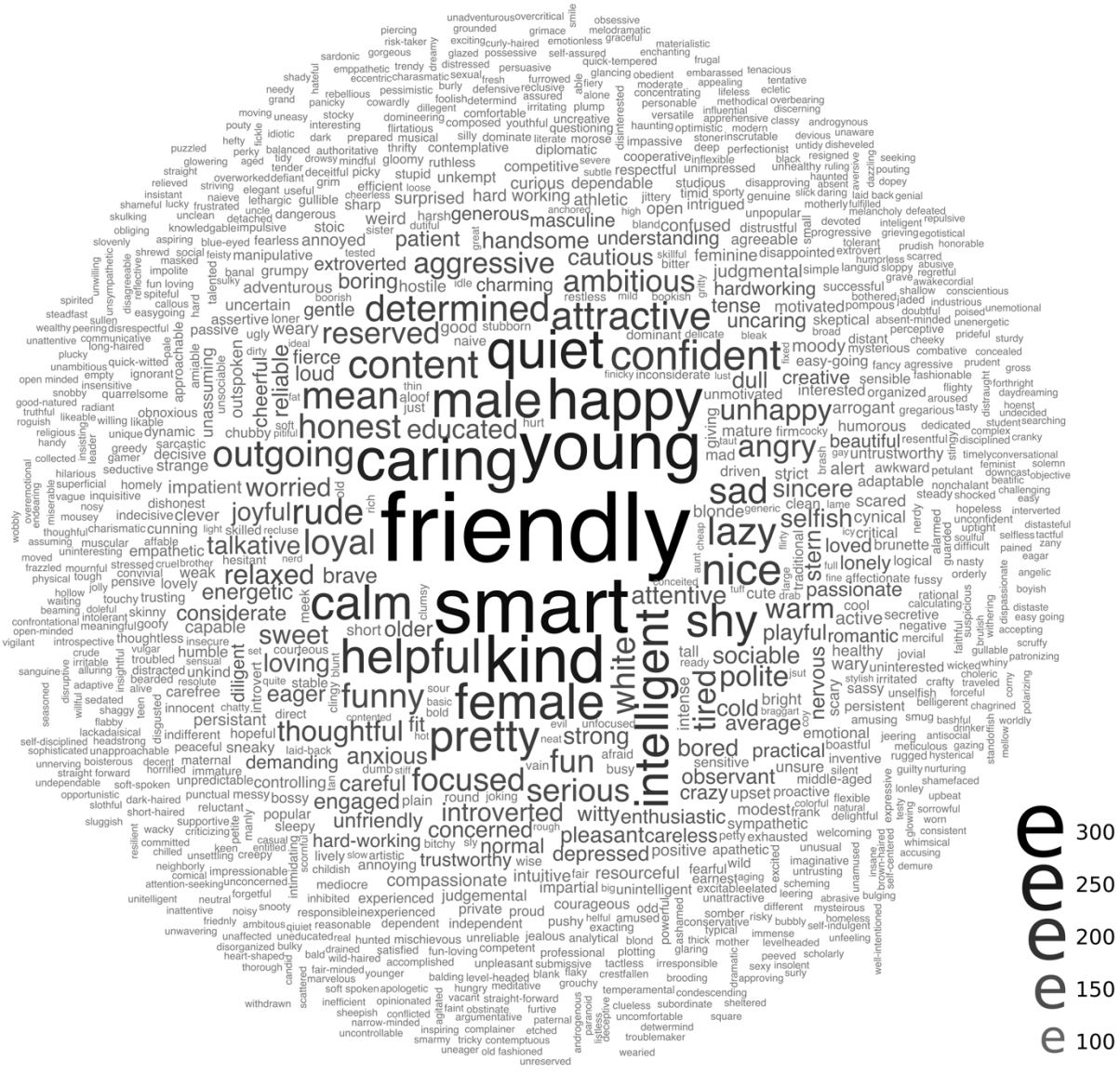
S5. Measures of Within-subject Test-retest Reliability and Between-subject Consensus (Study 2)

All participants at all locations rated a subset of twenty traits twice for all faces (see supplementary materials and methods section M6). Analyses of within-subject test-retest reliability identical to those in supplementary text section S2 were performed for each of the seven datasets ($l = 100$ pairs of ratings across faces per participant for ca. $n = 28$ participants at each location after data exclusion criteria A to C). We found acceptable within-subject test-retest reliabilities at all locations (except for the attributions of competent, religious, anxious, and critical in India [$rs = 0.18, 0.18, 0.19, 0.19$] and the attributions of anxious in Peru [$r = 0.19$]). As

hypothesized in our preregistration, across all locations, ratings of traits that were related to physical appearance had higher within-subject test-retest reliabilities (e.g., feminine, youthful, healthy, with mean *rs* = 0.74, 0.57, 0.51, respectively) than traits that were more abstract (e.g., critical, anxious, religious, with mean *r*s =0.31, 0.32, 0.33, respectively), corroborating findings from Study 1 (Fig. S3).

Assessment of between-subject consensus at each location used data from all participants within the same location (*l* = 100 ratings per participant for the 100 faces from ca. *n* = 28 participants per trait in each location after data exclusion criteria A to C). Assessment of cross-cultural consensus used data from all participants across seven locations. As hypothesized in our preregistration, traits that were related to physical appearance such as feminine, youthful, beautiful, and baby-faced showed high between-subject consensus in all seven locations and high cross-cultural consensus across all locations (all ICCs > 0.86). At the other extreme, some locations had trait ratings with near-zero consensus within that location (the ratings of compulsive in Gaza, prudish in India and Kenya, self-critical in Gaza and the Philippines). This stood in contrast to the findings from Study 1 where ICCs > 0.61 for all the one hundred traits (ca. *n* = 58 participants per trait who were white and located in the U.S.), and to the findings from North America (ca. *n* = 27 participants per trait; ICCs > 0.61 for all the eighty tested traits) and Latvia (ca. *n* = 28 participants per trait; ICCs > 0.50 for all the eighty tested traits).

A


friendly
smart
young
caring
quiet
confident
happy
male
kind
helpful
female
intelligent
calm
determined
attractive
content
aggressive
ambitious
pretty
focused
serious
introverted
outgoing
honest
reserved
mean
lazy
sad
nice
shy
polite
tired
cold
warm
sincere
selfish
stern
loyal
talkative
rude
relaxed
energetic
considerate
sweet
loving
eager
funny
thoughtful
careful
engaged
unfriendly
concerned
hard-working
trustworthy
normal
depressed
pleasant
careless
fun
strong
boring
hostile
charming
cautious
dependable
hard working
surprised
generous
masculine
handsome
understanding
patient
weird
annoyed
extroverted
unhappy
arrogant
beautiful
angry
mature
humorous
uncaring
tense
motivated
hardworking
successful
judgmental
dull
creative
sensible
unmotivated
interested
organized
unsympathetic
unwilling
disagreeable
reflective
talented
hard
unassuming
unsociable
outspoken
cheerful
reliable
fierce
loud
weary
passive
assertive
uncertain
adventurous
boorish
restless
bookish
grumpy
banal
naive
obnoxious
chubby
dynamic
likable
willing
unique
sarcastic
greedy
gamer
decisive
strange
seductive
inquisitive
vague
hilarious
superficial
homely
impatient
worried
joyful
clever
indecisive
dishonest
nosy
mousey
wobbly
overemotional
embarrassing
miserable
roguish
good-natured
truthful
likeable
radiant
religious
insisting
leader
collected
handy
steadfast
wealthy
unattentive
open minded
unambitious
snobby
spirited
callous
spiteful
fun loving
easygoing
disrespectful
communicative
long-haired
plucky
quick-witted
empty
ignorant
insensitive
quarrelsome
approachable
amiable
unobtrusive
unadventurous
overcritical
smile
obsessive
grounded
grimace
melodramatic
piercing
risk-taker
exciting
curly-haired
emotionless
graceful
materialistic
dreamy
sardonic
gorgeous
glazed
possessive
self-assured
enchanting
frugal
quick-tempered
empathetic
trendy
distressed
persuasive
tenacious
hateful
shady
eccentric
charasmatic
sexual
fresh
furrowed
able
glancing
obedient
embarrassed
tentative
needy
grand
rebellious
pessimistic
burly
defensive
reclusive
moderate
appealing
eclectic
panicky
cowardly
diligent
foolish
determind
assured
alone
fiery
concentrating
lifeless
moving
uneasy
stocky
domineering
comfortable
irritating
plump
disinterested
personable
methodical
overbearing
pouty
fickle
idiotic
dark
prepared
musical
silly
dominate
literate
morose
interesting
flirtatious
composed
youthful
questioning
uncreative
versatile
influential
discerning
hefty
puzzled
perky
balanced
authoritative
thrifty
contemplative
diplomatic
impassive
haunting
optimistic
modern
apprehensive
classy
androgynous
deep
perfectionist
stoner
inscrutable
devious
unaware
glowering
aged
tidy
drowsy
mindful
gloomy
ruthless
competitive
severe
cooperative
inflexible
black
resigned
seeking
straight
tender
deceitful
picky
stupid
unkempt
subtle
respectful
unimpressed
unhealthy
ruling
averse
dazzling
pouting
relieved
striving
overworked
defiant
grim
curious
studious
disapproving
haunted
absent
dopey
insistant
elegant
useful
efficient
loose
cheerless
sharp
timid
sporty
slick
daring
laid back
genial
shameful
lucky
frustrated
lethargic
gullible
anchored
high
open
intrigued
motherly
fulfilled
skulking
unclean
uncle
dangerous
detached
unpopular
melancholy
defeated
obliging
knowledgable
impulsive
stoic
sister
dutiful
great
bland
confused
distrustful
devoted
intelligent
repulsive
slovenly
aspiring
blue-eyed
fearless
tested
small
progressive
grieving
egotistical
shrewd
social
feisty
manipulative
tolerant
humorless
scarred
masked
impolite
banal
grumpy
extrovert
prudish
honorable
fun loving
spiteful
sulky
gritty
grave
abusive
callous
hard
easygoing
uncertain
gentle
boring
hostile
idle
charming
restless
mild
bookish
bothered
jaded
awakecordial
regretful
shallow
conscientious
steadfast
sullen
assertive
loner
determined
attractive
tense
motivated
pompous
doubtful
industrious
wealthy
peering
disrespectful
weary
reserved
good
stubborn
uncaring
skeptical
absent-minded
unemotional
unattentive
communicative
long-haired
naive
dominant
delicate
bleak
broad
perceptive
poised
unenergetic
plucky
quick-witted
ideal
dirty
fierce
loud
content
quiet
confident
moody
distant
prideful
unambitious
empty
ignorant
fixed
easy-going
fancy
agressive
prudent
sturdy
open minded
insensitive
quarrelsome
finicky
inconsiderate
lust
dull
creative
sensible
fashionable
gross
snobby
approachable
amiable
unsociable
outspoken
cheerful
reliable
fat
thin
aloof
just
mean
male
happy
unmotivated
interested
organized
aroused
forthright
daydreaming
good-natured
truthful
likeable
willing
likable
obnoxious
chubby
soft
pitiful
hurt
unhappy
arrogant
gregarious
tasty
disraught
honest
undecided
roguish
religious
insisting
leader
unique
dynamic
honest
educated
giving
laut
mature
firm
cocky
humorous
dedicated
student
searching
handy
sarcastic
mad
angry
beautiful
resentful
disciplined
complex
collected
gamer
decisive
strange
outgoing
caring
young
brash
gay
untrustworthy
stingy
cranky
hilarious
seductive
driven
strict
alert
awkward
petulant
feminist
solemn
timelyconversational
superficial
homely
impatient
worried
sad
sincere
adaptable
nonchalant
downcast
beatific
vague
inquisitive
old
rich
joyful
rude
blonde
generic
clean
lame
steady
shocked
challenging
dishonest
clever
indecisive
nosy
mousey
wobbly
overemotional
embarrassing
miserable
roguish
good-natured
truthful
likeable
willing
likable
obnoxious
chubby
soft
pitiful
hurt
unhappy
arrogant
gregarious
tasty
disraught
honest
undecided
handy
sarcastic
mad
angry
beautiful
resentful
disciplined
complex
collected
gamer
decisive
strange
outgoing
caring
young
brash
gay
untrustworthy
stingy
cranky
hilarious
seductive
driven
strict
alert
awkward
petulant
feminist
solemn
timelyconversational
superficial
homely
impatient
worried
sad
sincere
adaptable
nonchalant
downcast
beatific
vague
inquisitive
old
rich
joyful
rude
blonde
generic
clean
lame
steady
shocked
challenging
dishonest
nosy
indecisive
clever
mousey
wobbly
overemotional
embarrassing
miserable
roguish
good-natured
truthful
likeable
willing
likable
obnoxious
chubby
soft
pitiful
hurt
hopeless
interverted
unconfident
distasteful
thoughtful
charismatic
cunning
light
skilled
recluse
lazy
cheap
icy
critical
uptight
selfless
tactful
assuming
muscular
affable
talkative
loyal
empathetic
hesitant
nerd
lonely
logical
brunette
difficult
pained
zany
moved
frazzled
mourful
stressed
cruel
brother
relaxed
brave
weak
convivial
full
fine
affectionate
fussy
orderly
eagar
physical
tough
jolly
pensive
lovely
energetic
meek
clumsy
nice
shy
attentive
conceited
tuff
cute
drab
large
traditional
passionate
angelic
hollow
touchy
trusting
waiting
beaming
doleful
intolerant
meaningful
goofy
skinny
considerate
capable
sweet
short
older
calculating
boyish
confrontational
open-minded
vigilant
introspective
insecure
thoughtless
humble
warm
cool
active
secretive
negative
distaste
easy going
crude
irritable
sanguine
alluring
bearded
resolute
troubled
sensual
loving
diligent
introvert
set
courteous
playful
romantic
merciful
accepting
scruffy
adaptive
alive
sedated
shaggy
teen
carefree
eager
stable
quite
funny
female
white
intelligent
tall
ready
sociable
polite
jsut
nervous
wary
uninterested
healthy
jovial
faithful
suspicious
patronizing
seasoned
disruptive
willful
innocent
chatty
direct
clingy
basic
sour
intense
tired
cold
bright
scary
sassy
unselfish
forceful
choleric
whiny
corny
polarizing
lackadaisical
indifferent
hopeful
persistant
disgusted
contented
fit
hot
evil
unfocused
average
braggart
coy
persistent
belligerent
chagrined
amusing
smug
bashful
drinker
worldly
self-disciplined
headstrong
sophisticated
unapproachable
peaceful
sneaky
laid-back
thoughtful
pretty
strong
neat
emotional
jeering
antisocial
standoffish
mellow
unnerving
boisterous
decent
maternal
demanding
anxious
fun
afraid
vain
boastful
meticulous
gazing
straight forward
horrified
immature
dumb
stiff
bored
practical
inventive
rugged
hysterical
undependable
soft-spoken
unpredictable
controlling
careful
focused
serious
busy
sensitive
unsure
silent
guilty
nurturing
shamefaced
opportunistic
dark-haired
punctual
messy
bossy
engaged
plain
round
joking
observant
middle-aged
crazy
upset
proactive
flexible
upbeat
slothful
short-haired
reluctant
popular
introverted
witty
enthusiastic
colorful
natural
lonley
sorrowful
sluggish
supportive
petite
mainly
sleepy
unfriendly
concerned
rough
modest
frank
delightful
expressive
worn
wacky
criticizing
casual
pleasant
careless
petty
exhausted
sympathetic
welcoming
consistent
resilient
committed
chilled
entitled
keen
hard-working
bitchy
sly
normal
depressed
positive
apathetic
unusual
whimsical
unsettling
creepy
lively
slow
artistic
trustworthy
wise
excited
imaginative
accusing
neighborly
impressionable
childish
annoying
intuitive
fair
resourceful
fearful
wild
demure
comical
unconcerned
mediocre
compassionate
earnest
aging
untrusting
attention-seeking
inhibited
experienced
judgemental
impartial
big
unintelligent
excitable
elated
scheming
unattractive
leering
abrasive
bulging
unitelligent
neutral
forgetful
inattentive
noisy
snooty
responsible
inexperienced
private
proud
courageous
odd
somber
risky
different
mysteirous
friendly
ambitous
quiet
reasonable
dependent
independent
pushy
helful
amused
exacting
conservative
homeless
bubbly
self-indulgent
unwavering
unaffected
uneducated
real
hunted
mischievous
unreliable
jealous
analytical
blond
powerful
typical
immense
unfeeling
disorganized
bulky
drained
satisfied
fun-loving
competent
professional
plotting
glaring
peeved
scholarly
heart-shaped
bald
wild-haired
accomplished
unpleasant
submissive
tactless
irresponsible
sexy
insolent
thorough
fair-minded
younger
balding
level-headed
blank
flaky
crestfallen
brooding
approving
surly
candid
marvelous
soft spoken
apologetic
hungry
meditative
grouchy
temperamental
condescending
withdrawn
inefficient
opinionated
vacant
straight-forward
clueless
subordinate
sheltered
sheepish
conflicted
faint
obstinate
furtive
uncomfortable
square
narrow-minded
inspiring
complainer
paternal
detwermind
uncontrollable
smarmy
tricky
contemptuous
etched
troublemaker
uneager
old fashioned
wearied
unreserved
300
250
200
150
100
50

B

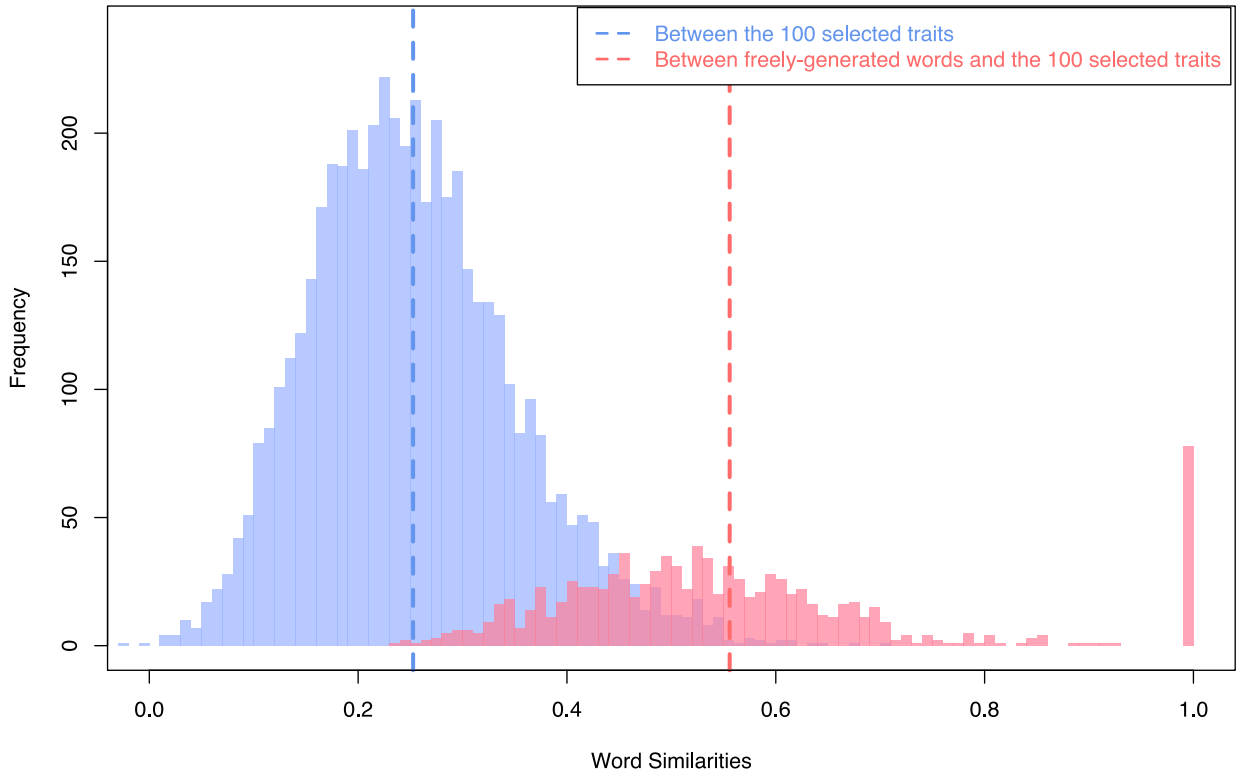


Fig. S1. Freely-generated words and their similarities with the 100 selected traits. (**A**) Word cloud of 973 freely-generated words. All words that appeared at least twice are shown (words appeared only once were mostly misspelled words or words not included in the FastText vocabulary (*61*) and were therefore excluded, as preregistered). The scale indicates frequency (ranged from 2 to 306 times). (**B**) Distributions of word similarities. The similarity between two words was assessed with the cosine distance between the 300-feature vectors of the two words. The blue histogram plots the pairwise similarities among the 100 selected traits. The red histogram plots the similarities between the freely-generated words and their closest counterparts in our selected traits. Dashed lines indicate mean similarities.

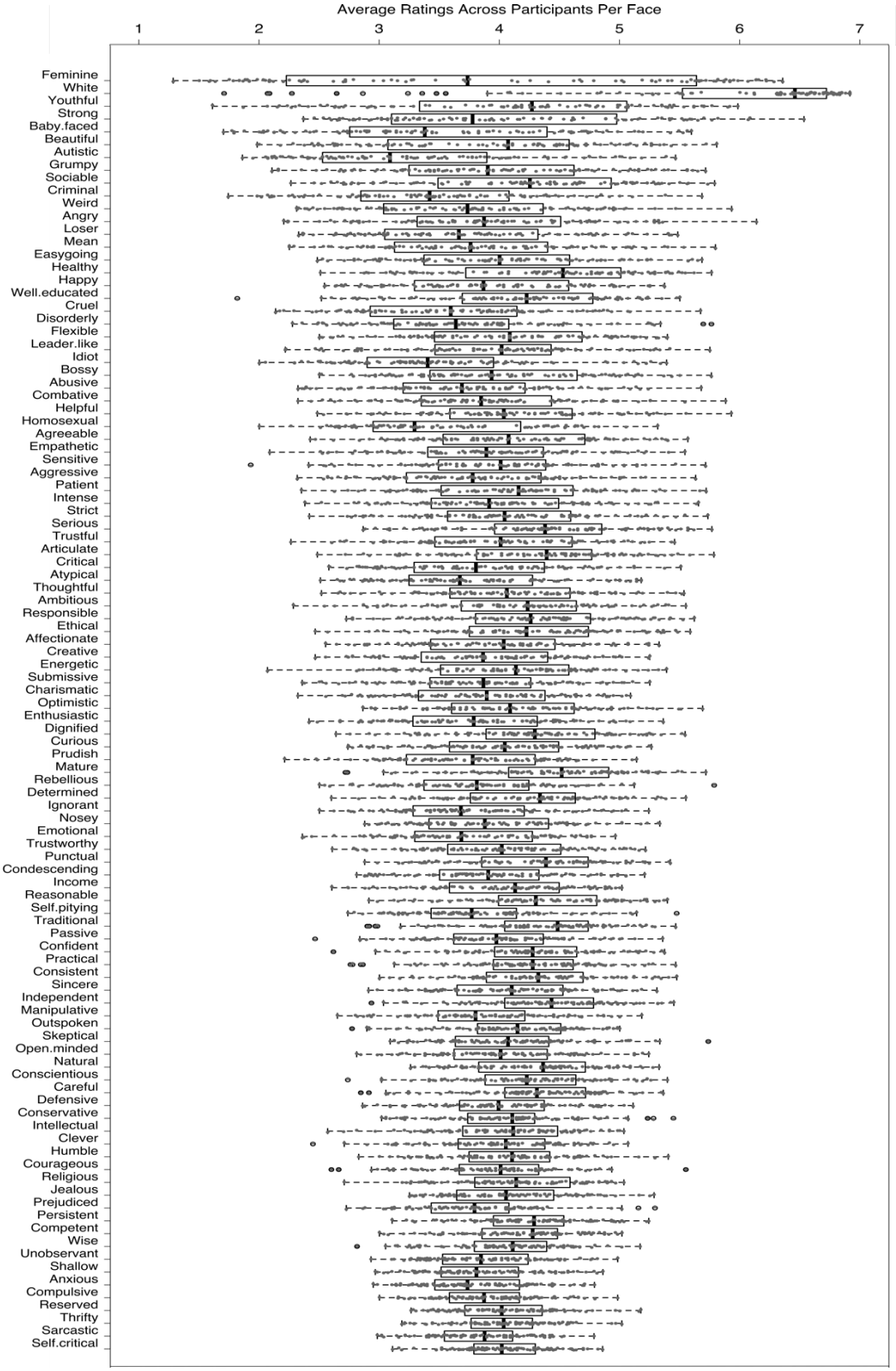

Average Ratings Across Participants Per Face
1
2
3
4
5
6
7
Feminine
White
Youthful
Strong
Baby.faced
Beautiful
Autistic
Grumpy
Sociable
Criminal
Weird
Angry
Loser
Mean
Easygoing
Healthy
Happy
Well.educated
Cruel
Disorderly
Flexible
Leader.like
Idiot
Bossy
Abusive
Combative
Helpful
Homosexual
Agreeable
Empathetic
Sensitive
Aggressive
Patient
Intense
Strict
Serious
Trustful
Articulate
Critical
Atypical
Thoughtful
Ambitious
Responsible
Ethical
Affectionate
Creative
Energetic
Submissive
Charismatic
Optimistic
Enthusiastic
Dignified
Curious
Prudish
Mature
Rebellious
Determined
Ignorant
Nosey
Emotional
Trustworthy
Punctual
Condescending
Income
Reasonable
Self.pitying
Traditional
Passive
Confident
Practical
Consistent
Sincere
Independent
Manipulative
Outspoken
Skeptical
Open.minded
Natural
Conscientious
Careful
Defensive
Conservative
Intellectual
Clever
Humble
Courageous
Religious
Jealous
Prejudiced
Persistent
Competent
Wise
Unobservant
Shallow
Anxious
Compulsive
Reserved
Thrifty
Sarcastic
Self.critical

Fig. S2. Distributions of average ratings per face for the 100 traits.
Each row plots the average ratings across participants for the 100 faces on a trait (grey dots), with the boxplot indicating the median (line in the box), the first quartile (left edge of the box), the third quartile (right edge of the box), the minimum excluding outliers (leftmost line), the maximum excluding outliers (rightmost line), and the outliers that are more extreme than 3/2 times of the first or third quartiles (open dots).

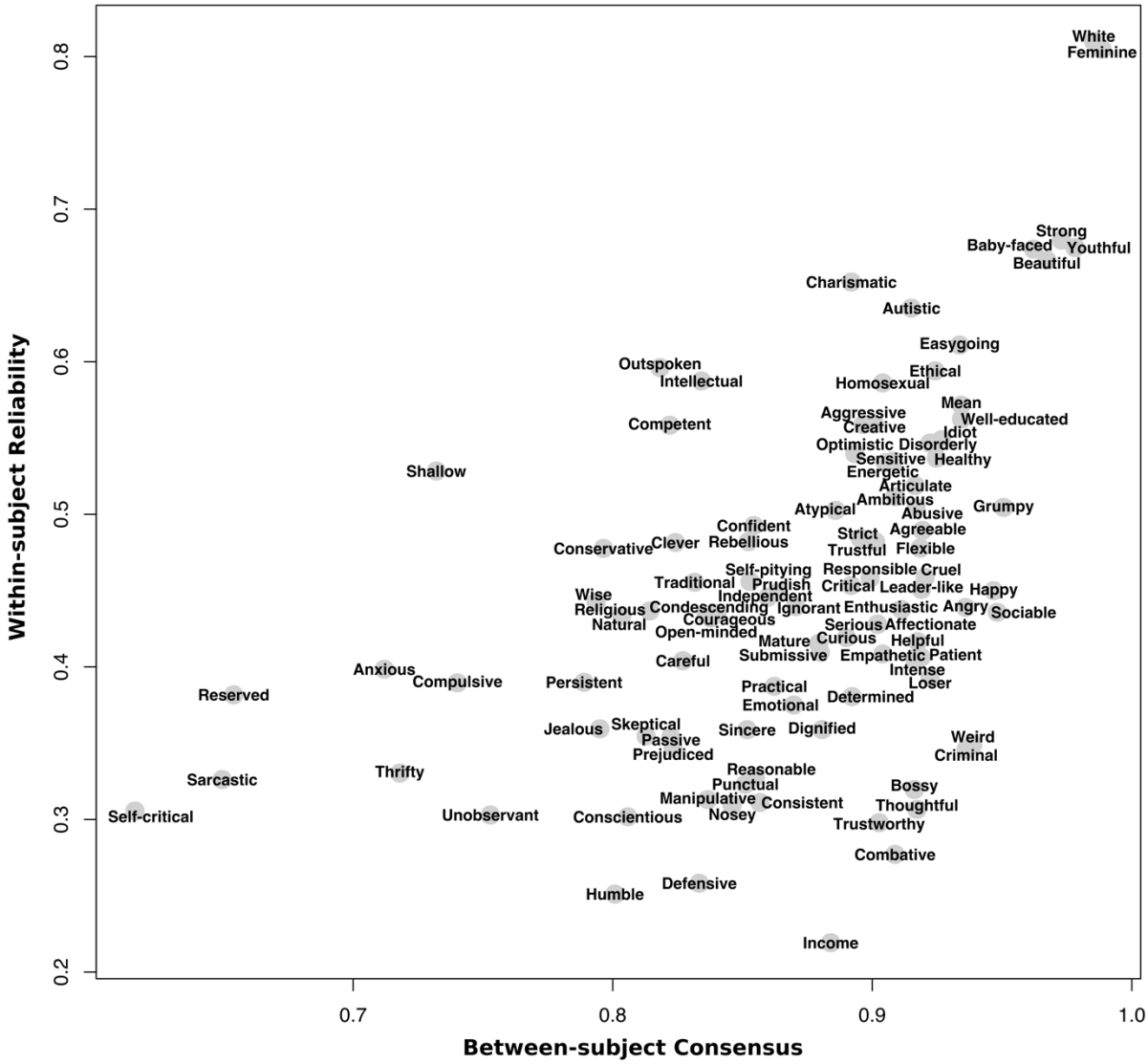


Fig. S3. Within-subject test-retest reliability and between-subject consensus for 100 traits. The vertical axis indicates the within-subject test-retest reliability assessed with linear mixed-effect modeling (regression coefficients). The horizontal axis indicates the between-subject consensus assessed with intraclass correlation coefficients (ICCs).

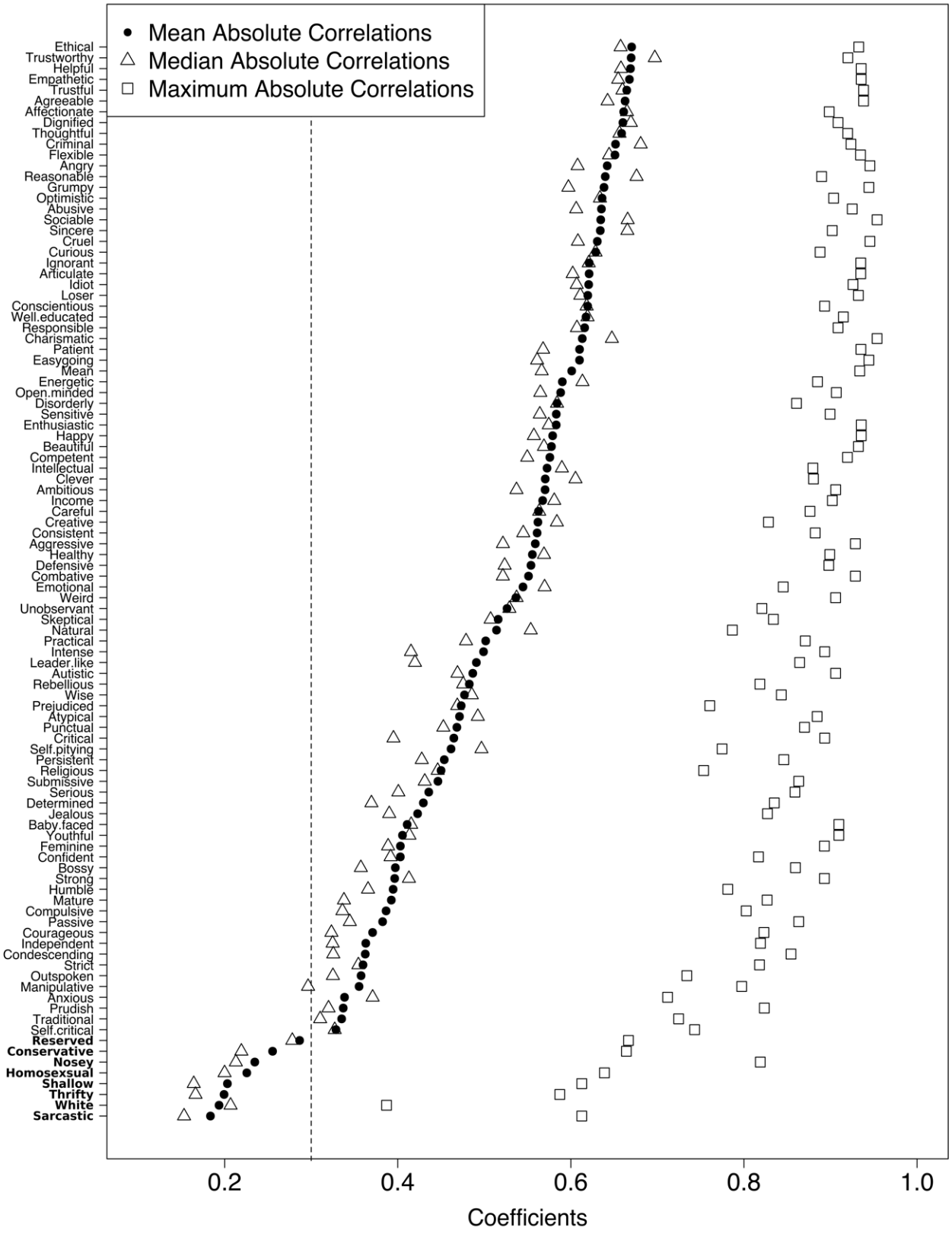

Mean Absolute Correlations
Median Absolute Correlations
Maximum Absolute Correlations
Ethical
Trustworthy
Helpful
Empathetic
Trustful
Agreeable
Affectionate
Dignified
Thoughtful
Criminal
Flexible
Angry
Reasonable
Grumpy
Optimistic
Abusive
Sociable
Sincere
Cruel
Curious
Ignorant
Articulate
Idiot
Loser
Conscientious
Well.educated
Responsible
Charismatic
Patient
Easygoing
Mean
Energetic
Open.minded
Disorderly
Sensitive
Enthusiastic
Happy
Beautiful
Competent
Intellectual
Clever
Ambitious
Income
Careful
Creative
Consistent
Aggressive
Healthy
Defensive
Combative
Emotional
Weird
Unobservant
Skeptical
Natural
Practical
Intense
Leader.like
Autistic
Rebellious
Wise
Prejudiced
Atypical
Punctual
Critical
Self.pitying
Persistent
Religious
Submissive
Serious
Determined
Jealous
Baby.faced
Youthful
Feminine
Confident
Bossy
Strong
Humble
Mature
Compulsive
Passive
Courageous
Independent
Condescending
Strict
Outspoken
Manipulative
Anxious
Prudish
Traditional
Self.critical
Reserved
Conservative
Nosey
Homosexsual
Shallow
Thrifty
White
Sarcastic
0.2
0.4
0.6
0.8
1.0
Coefficients

Fig. S4. Factorizability assessment of the 100 trait ratings.

The vertical axis indicates traits and the horizontal axis indicates correlation coefficients. Each symbol plots the mean (dot), median (triangle), and maximum (square) absolute correlations a trait has with all the other ninety-nine traits (across faces, averaged over participants). The vertical dashed line indicates $r = 0.30$, which describes an inflection point in the curve of mean absolute correlations. The eight traits at the bottom (in bold) were excluded from EFA because of their low average correlations with all other traits (i.e. low factorability). Including these eight traits do not change the estimation of the optimal number of factors and the results from EFA.

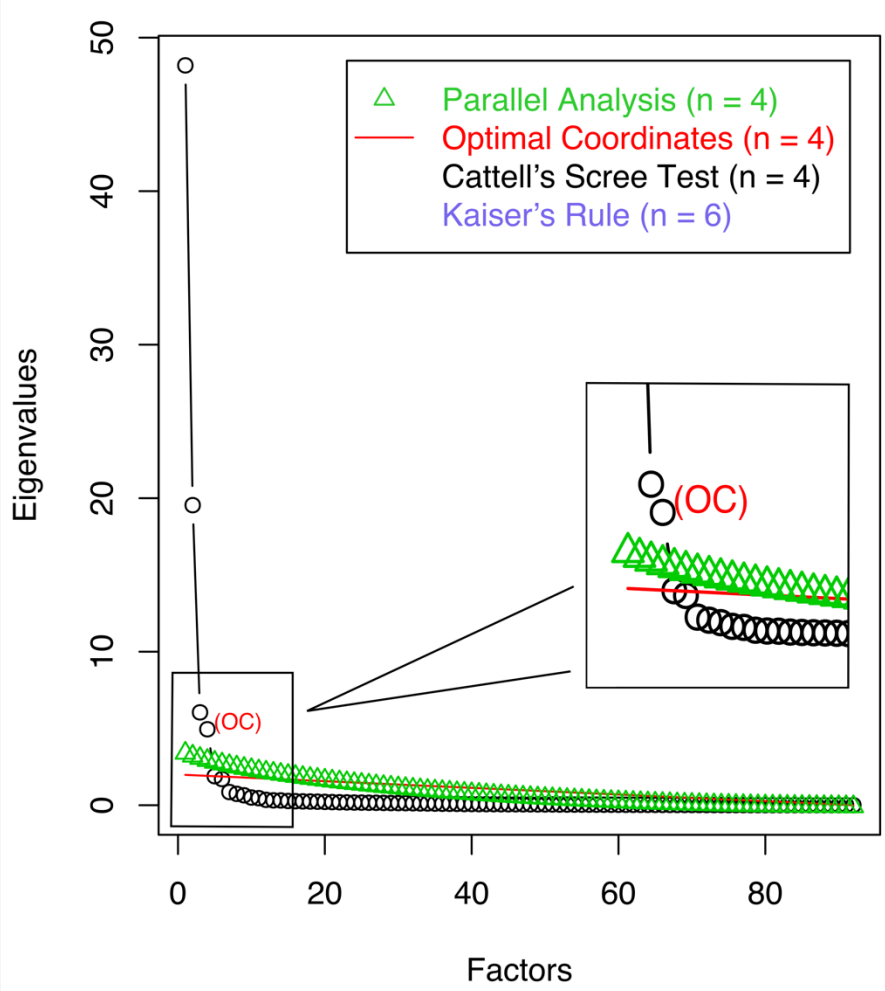


Fig. S5. Scree plot of 92 trait attributions from faces.

The horizontal axis indicates factors. The vertical axis indicates the fraction of total common variance in the data as explained by each factor. Circles plot the eigenvalues of the original data, ordered from the largest to the smallest. Triangles plot the 95th percentile of the eigenvalues of the simulated data from parallel analysis. The optimal number of factors to retain as recommended by each of the four methods is shown. Parallel analysis retains factors with eigenvalues (circles) greater than those (triangles) from the simulated data (see the close-up image for a clearer comparison). Cattell's scree test retains factors to the left of the point from which the plotted ordered eigenvalues could be approximated with a straight line (i.e., "above the elbow"). The optimal coordinates index provides a non-graphical solution to Cattell's scree test based on linear extrapolation. Kaiser's rule retains factors with eigenvalues that are greater than one.

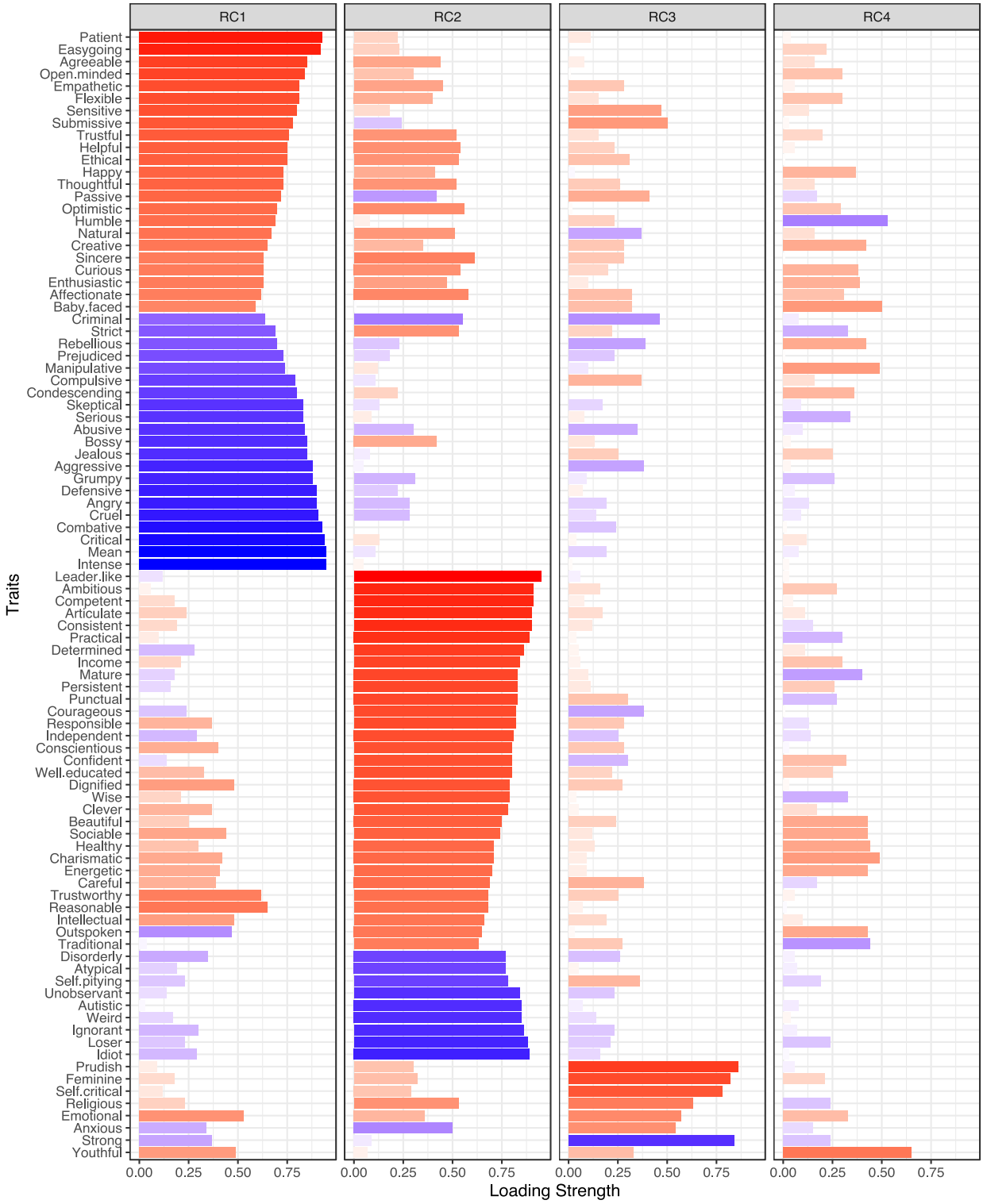
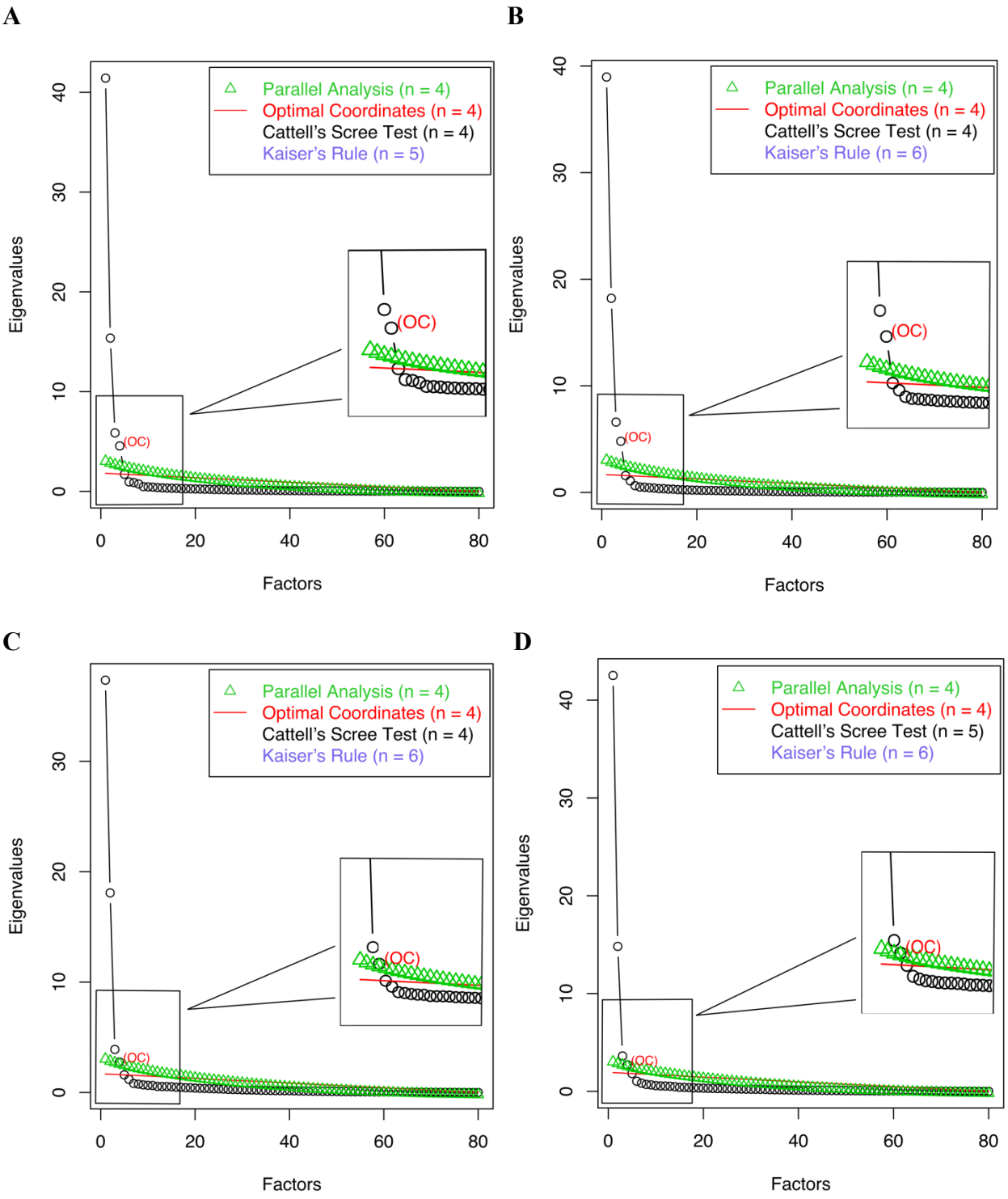


Fig. S6. Standard loadings of the first four rotated principal components.
Columns plot the strength of the loadings (x-axis, absolute value) on the first four varimax rotated principal components across all 92 traits (y-axis). Colors indicate the sign of the loading (red for positive and blue for negative); more saturated colors for higher absolute values.

Fig. S7. Predicting trait attributions using different dimensional frameworks.
Each row plots adjusted R-squared from regressing attributions of a trait on the predictors from three different frameworks. The framework from (*8*) offers two predictors (trustworthy and submissive) [yellow dots]. The framework from (*35*) offers four predictors (trustworthy, submissive, youthful, and beautiful) [blue triangles]. The framework from our present study offers four predictors (agreeable, leader-like, feminine, and youthful) [red squares].


A
Parallel Analysis (n = 4)
Optimal Coordinates (n = 4)
Cattell's Scree Test (n = 4)
Kaiser's Rule (n = 5)
(OC)
Eigenvalues
Factors
B
Parallel Analysis (n = 4)
Optimal Coordinates (n = 4)
Cattell's Scree Test (n = 4)
Kaiser's Rule (n = 6)
(OC)
Eigenvalues
Factors
C
Parallel Analysis (n = 4)
Optimal Coordinates (n = 4)
Cattell's Scree Test (n = 4)
Kaiser's Rule (n = 6)
(OC)
Eigenvalues
Factors
D
Parallel Analysis (n = 4)
Optimal Coordinates (n = 4)
Cattell's Scree Test (n = 5)
Kaiser's Rule (n = 6)
(OC)
Eigenvalues
Factors

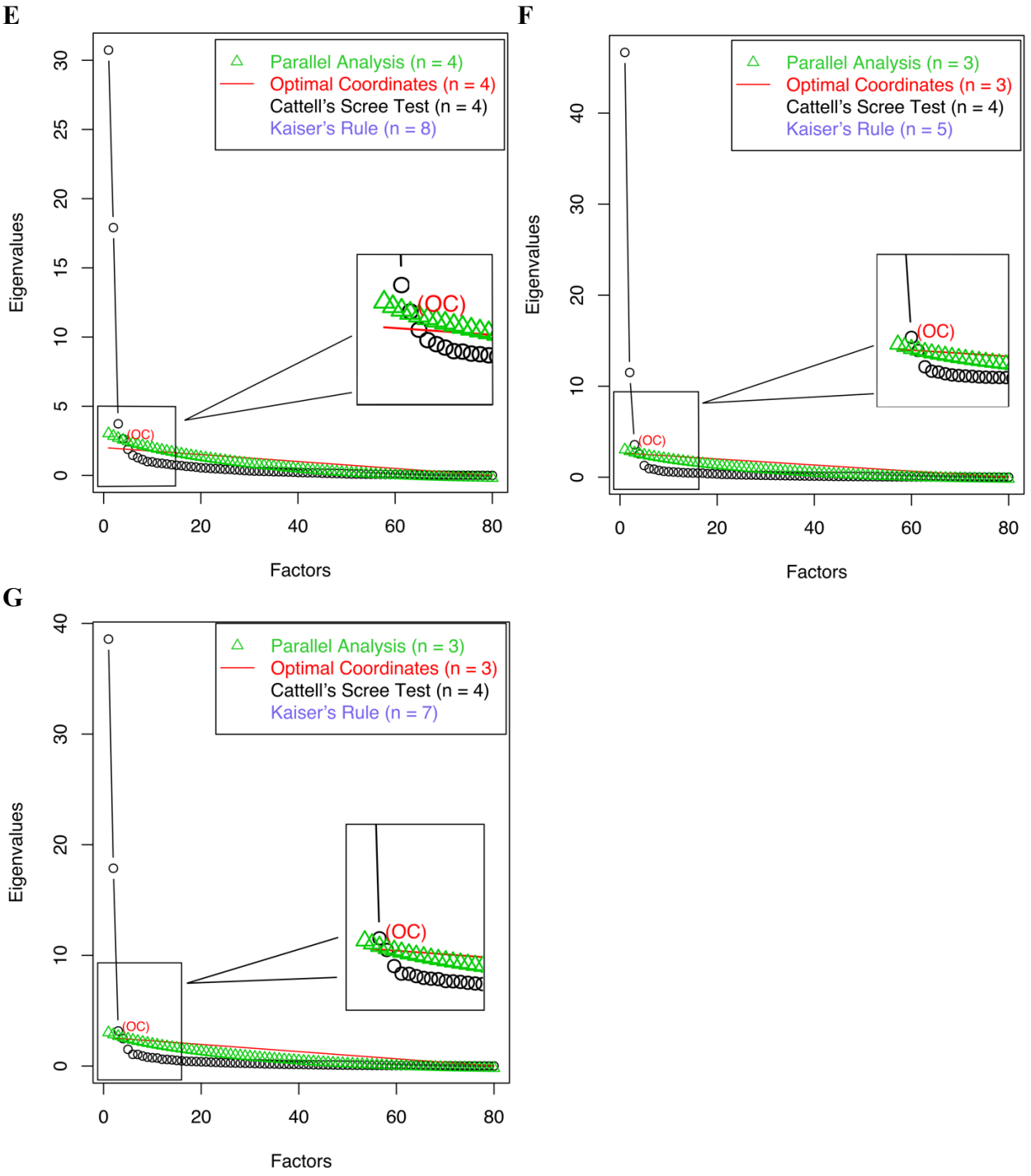
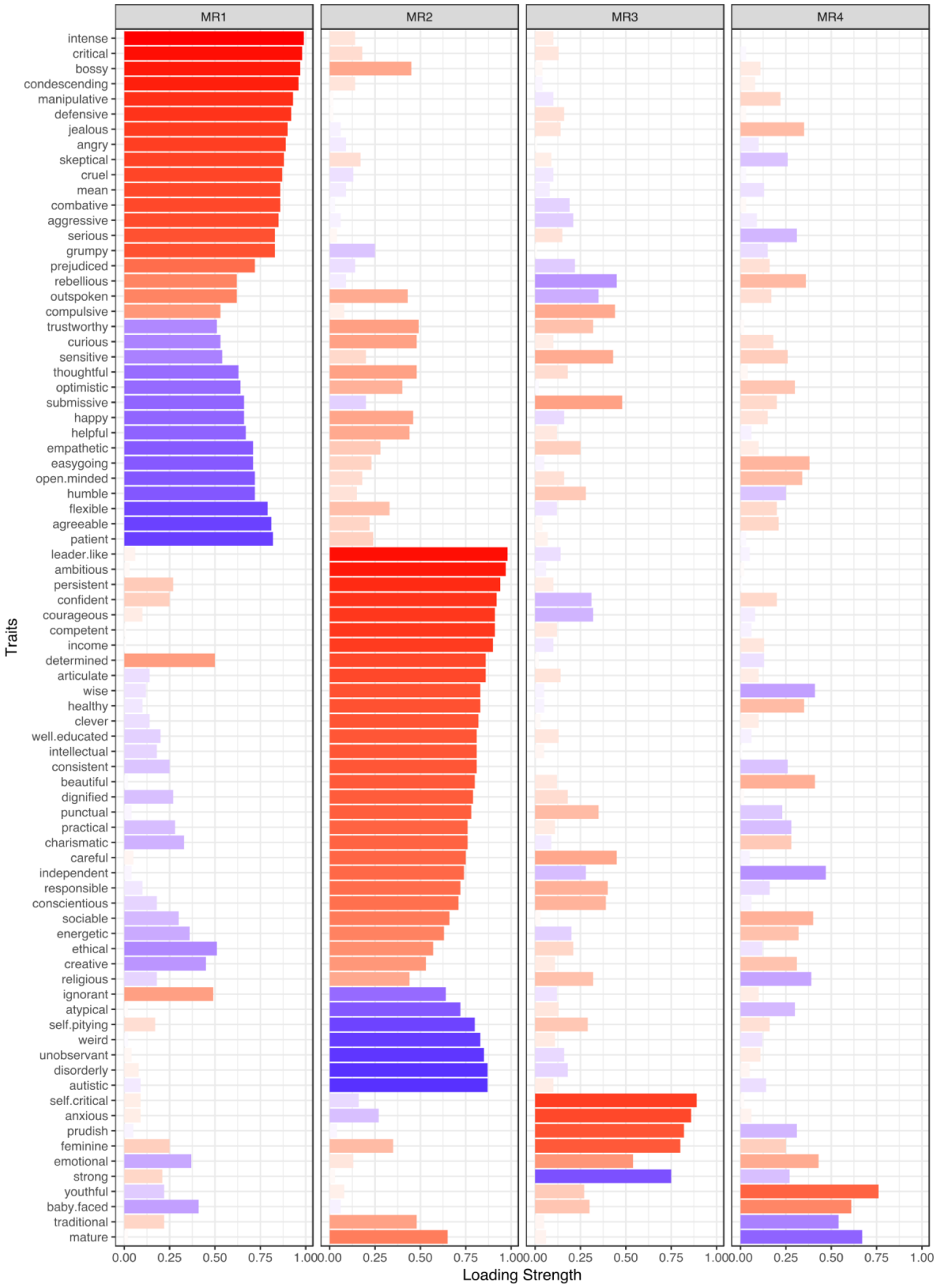


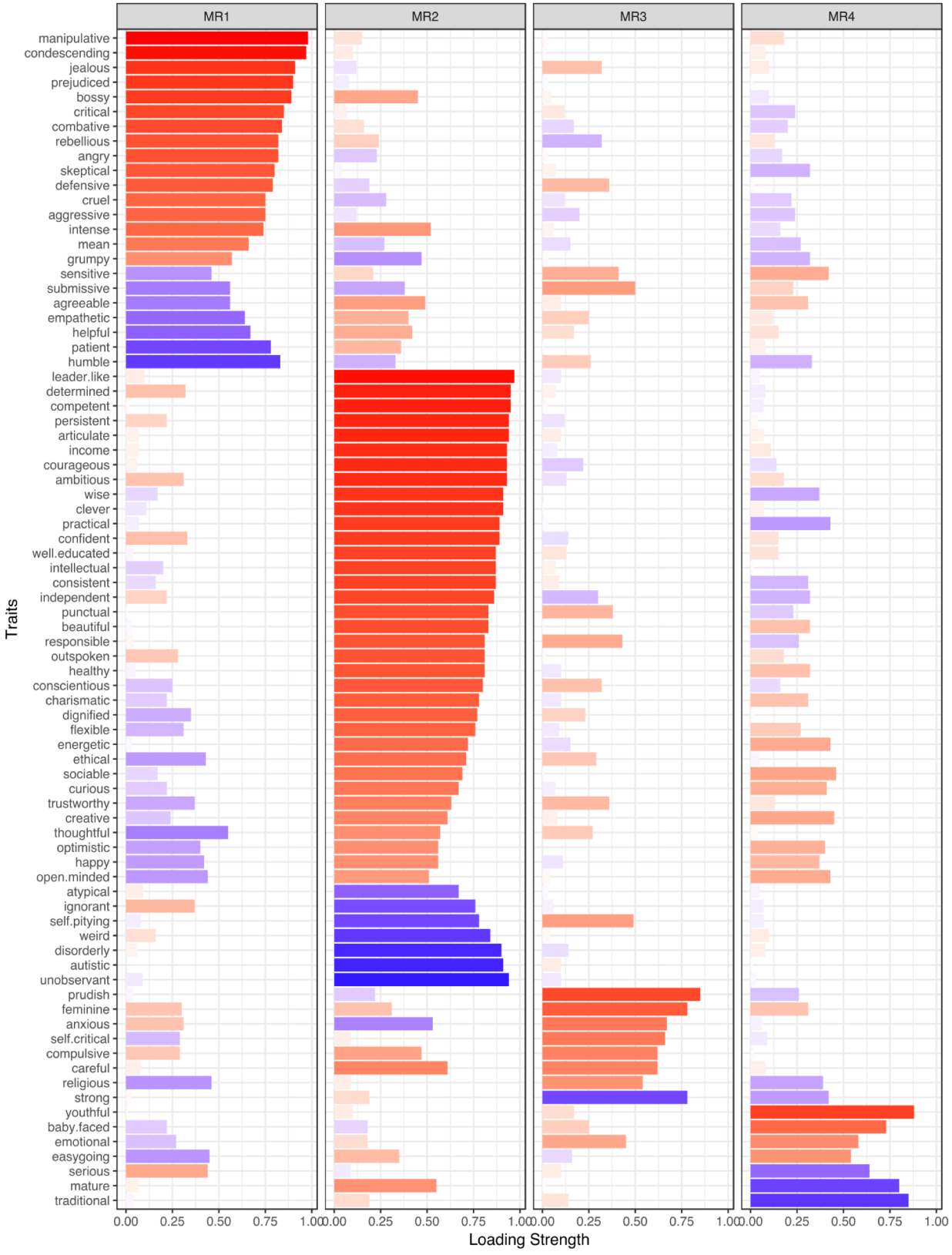
Fig. S8. Scree plot of 80 trait attributions from faces across seven samples.
The seven panels present scree plots for samples from North America (**A**), Latvia (**B**), Peru (**C**), the Philippines (**D**), India (**E**), Kenya (**F**), and Gaza (**G**). The horizontal axis indicates factors. The vertical axis indicates the fraction of common variance in the data as explained by each factor. Circles plot eigenvalues of the original data, ordered from the largest to the smallest. Triangles plot the 95th percentile of the eigenvalues of the simulated data from parallel analysis.

The optimal number of factors to retain as recommended by each of the four methods is shown. Parallel analysis retains factors with eigenvalues (circles) greater than those (triangles) from the simulated data (see the close-up image for a clearer comparison). Cattell's scree test retains factors to the left of the point from which the plotted ordered eigenvalues could be approximated with a straight line (i.e., "above the elbow"). The optimal coordinates index provides a non-graphical solution to Cattell's scree test based on linear extrapolation. Kaiser's rule retains factors with eigenvalues that are greater than one.

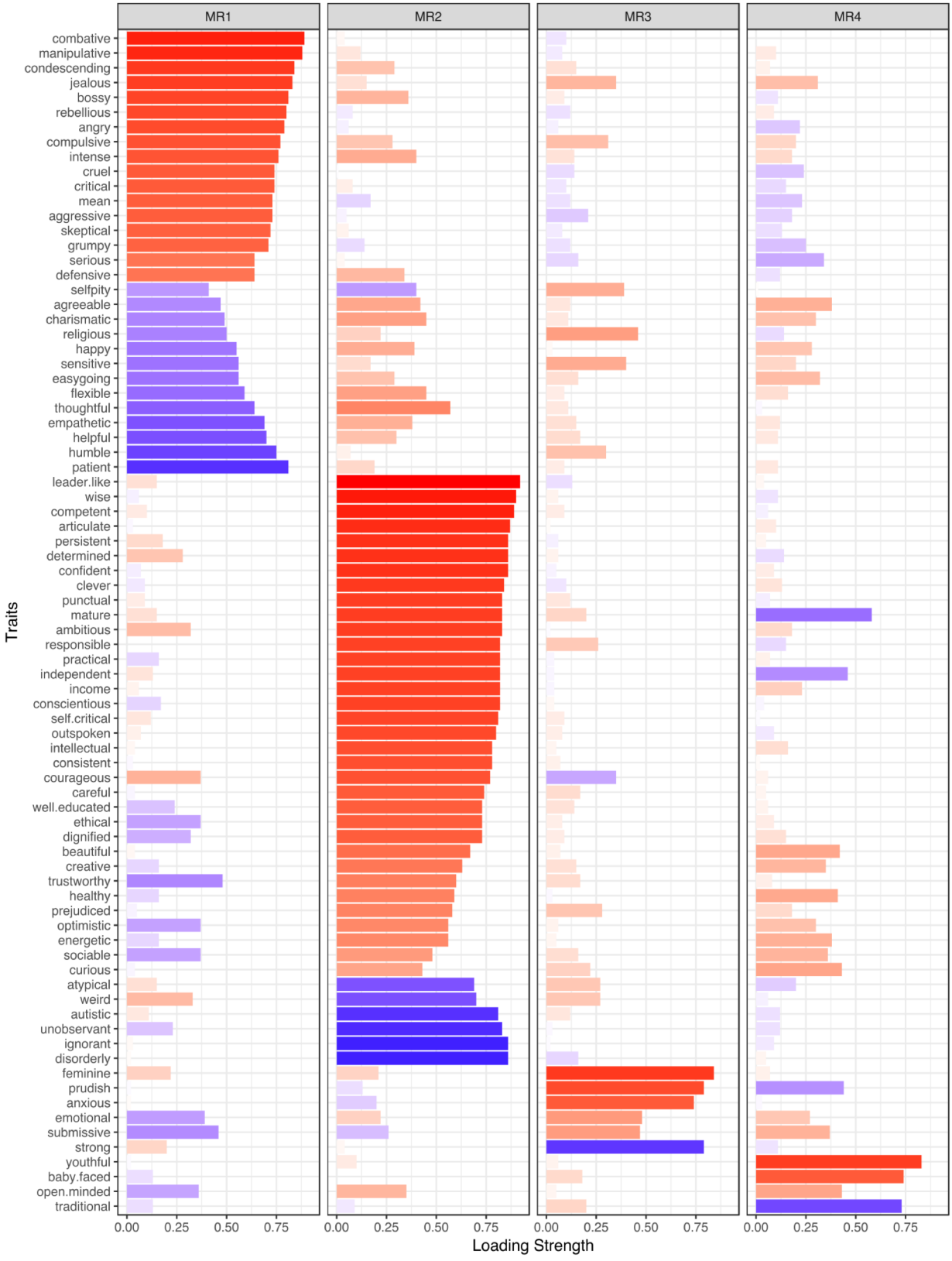
A


MR1
MR2
MR3
MR4
Traits
intense
critical
bossy
condescending
manipulative
defensive
jealous
angry
skeptical
cruel
mean
combative
aggressive
serious
grumpy
prejudiced
rebellious
outspoken
compulsive
trustworthy
curious
sensitive
thoughtful
optimistic
submissive
happy
helpful
empathetic
easygoing
open.minded
humble
flexible
agreeable
patient
leader.like
ambitious
persistent
confident
courageous
competent
income
determined
articulate
wise
healthy
clever
well.educated
intellectual
consistent
beautiful
dignified
punctual
practical
charismatic
careful
independent
responsible
conscientious
sociable
energetic
ethical
creative
religious
ignorant
atypical
self.pitying
weird
unobservant
disorderly
autistic
self.critical
anxious
prudish
feminine
emotional
strong
youthful
baby.faced
traditional
mature
0.00
0.25
0.50
0.75
1.00
Loading Strength

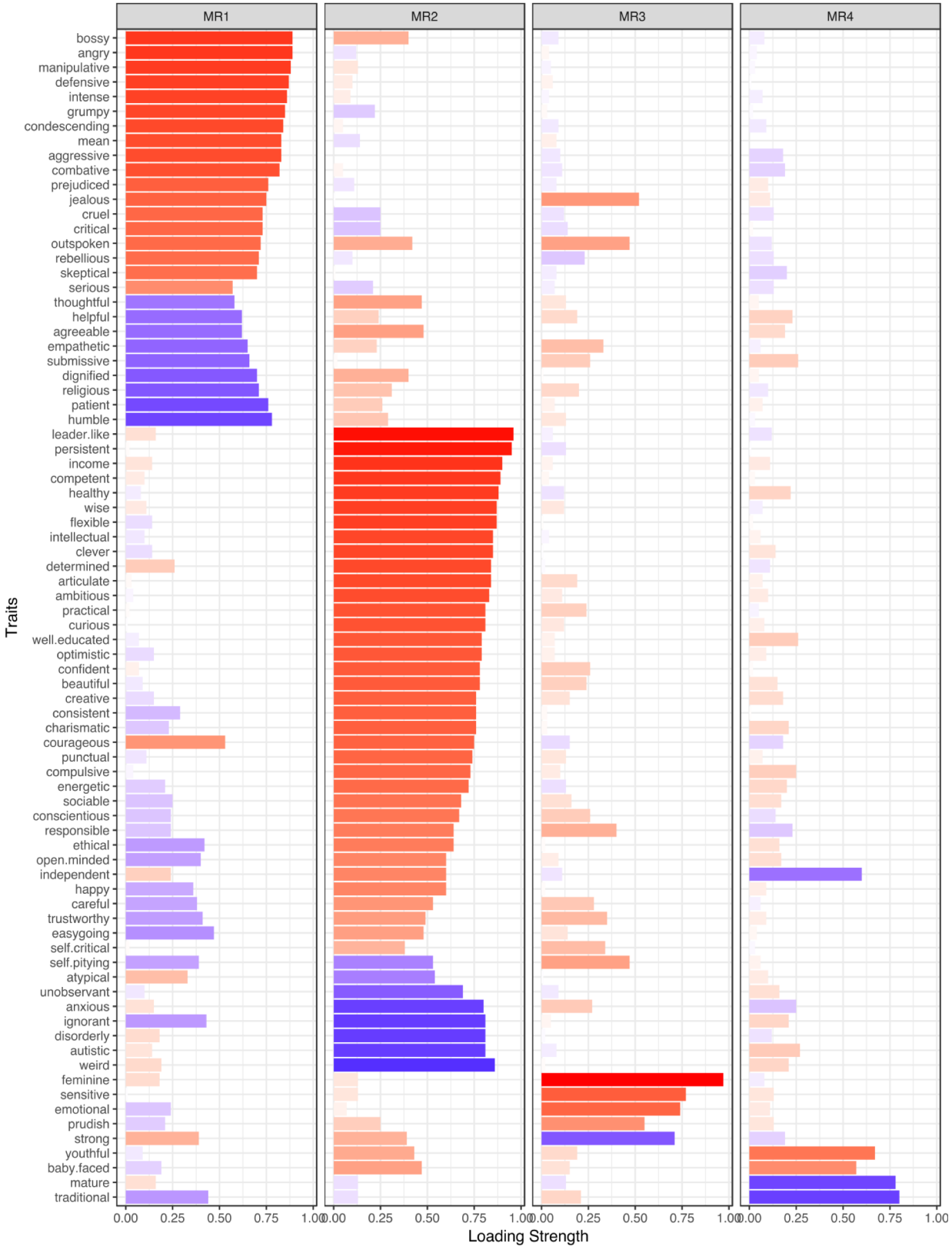
B


MR1
MR2
MR3
MR4
Traits
manipulative
condescending
jealous
prejudiced
bossy
critical
combative
rebellious
angry
skeptical
defensive
cruel
aggressive
intense
mean
grumpy
sensitive
submissive
agreeable
empathetic
helpful
patient
humble
leader.like
determined
competent
persistent
articulate
income
courageous
ambitious
wise
clever
practical
confident
well.educated
intellectual
consistent
independent
punctual
beautiful
responsible
outspoken
healthy
conscientious
charismatic
dignified
flexible
energetic
ethical
sociable
curious
trustworthy
creative
thoughtful
optimistic
happy
open.minded
atypical
ignorant
self.pitying
weird
disorderly
autistic
unobservant
prudish
feminine
anxious
self.critical
compulsive
careful
religious
strong
youthful
baby.faced
emotional
easygoing
serious
mature
traditional
0.00
0.25
0.50
0.75
1.00
Loading Strength

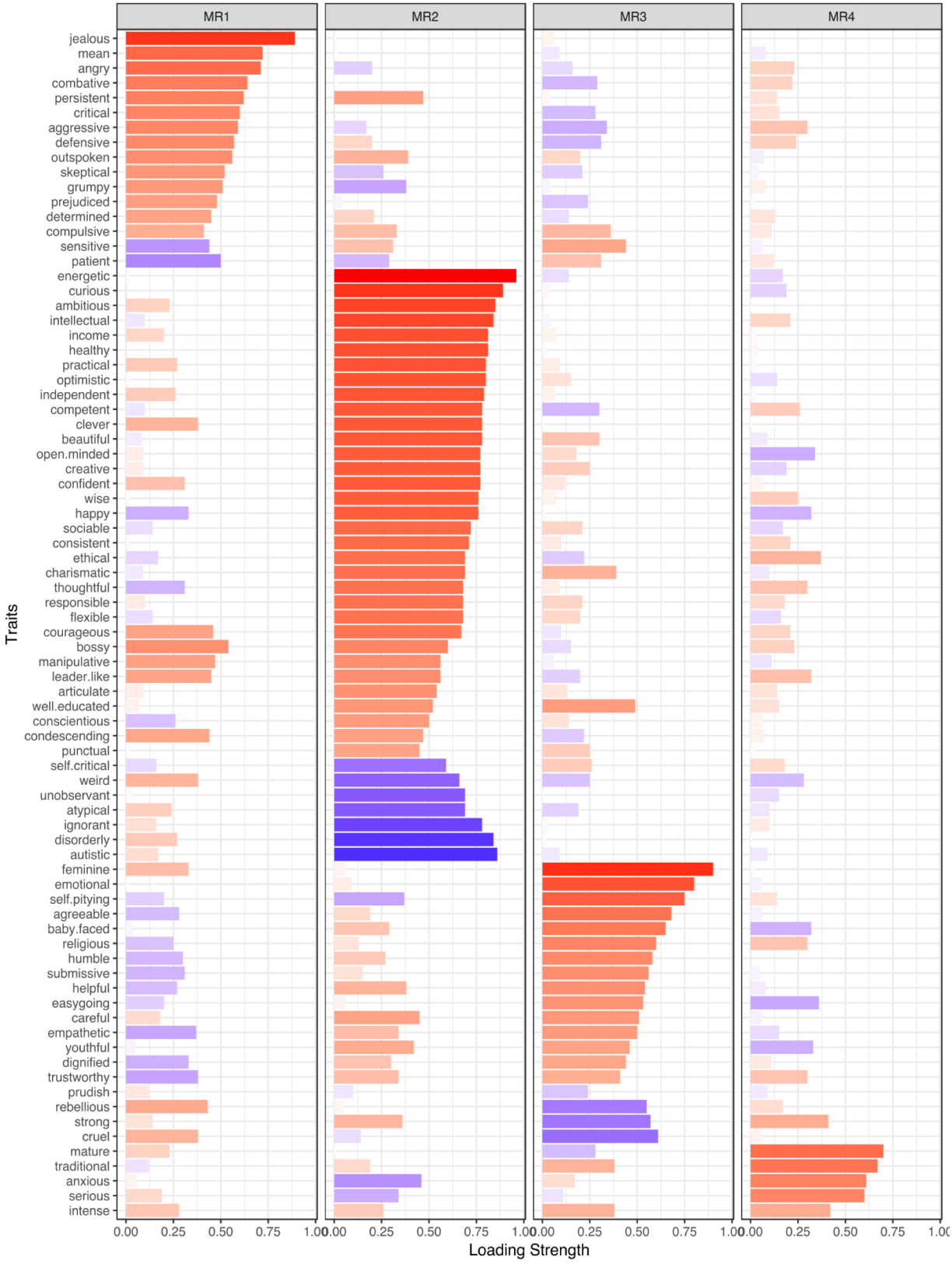
C


MR1
MR2
MR3
MR4
Traits
combative
manipulative
condescending
jealous
bossy
rebellious
angry
compulsive
intense
cruel
critical
mean
aggressive
skeptical
grumpy
serious
defensive
selfpity
agreeable
charismatic
religious
happy
sensitive
easygoing
flexible
thoughtful
empathetic
helpful
humble
patient
leader.like
wise
competent
articulate
persistent
determined
confident
clever
punctual
mature
ambitious
responsible
practical
independent
income
conscientious
self.critical
outspoken
intellectual
consistent
courageous
careful
well.educated
ethical
dignified
beautiful
creative
trustworthy
healthy
prejudiced
optimistic
energetic
sociable
curious
atypical
weird
autistic
unobservant
ignorant
disorderly
feminine
prudish
anxious
emotional
submissive
strong
youthful
baby.faced
open.minded
traditional
0.00
0.25
0.50
0.75
Loading Strength

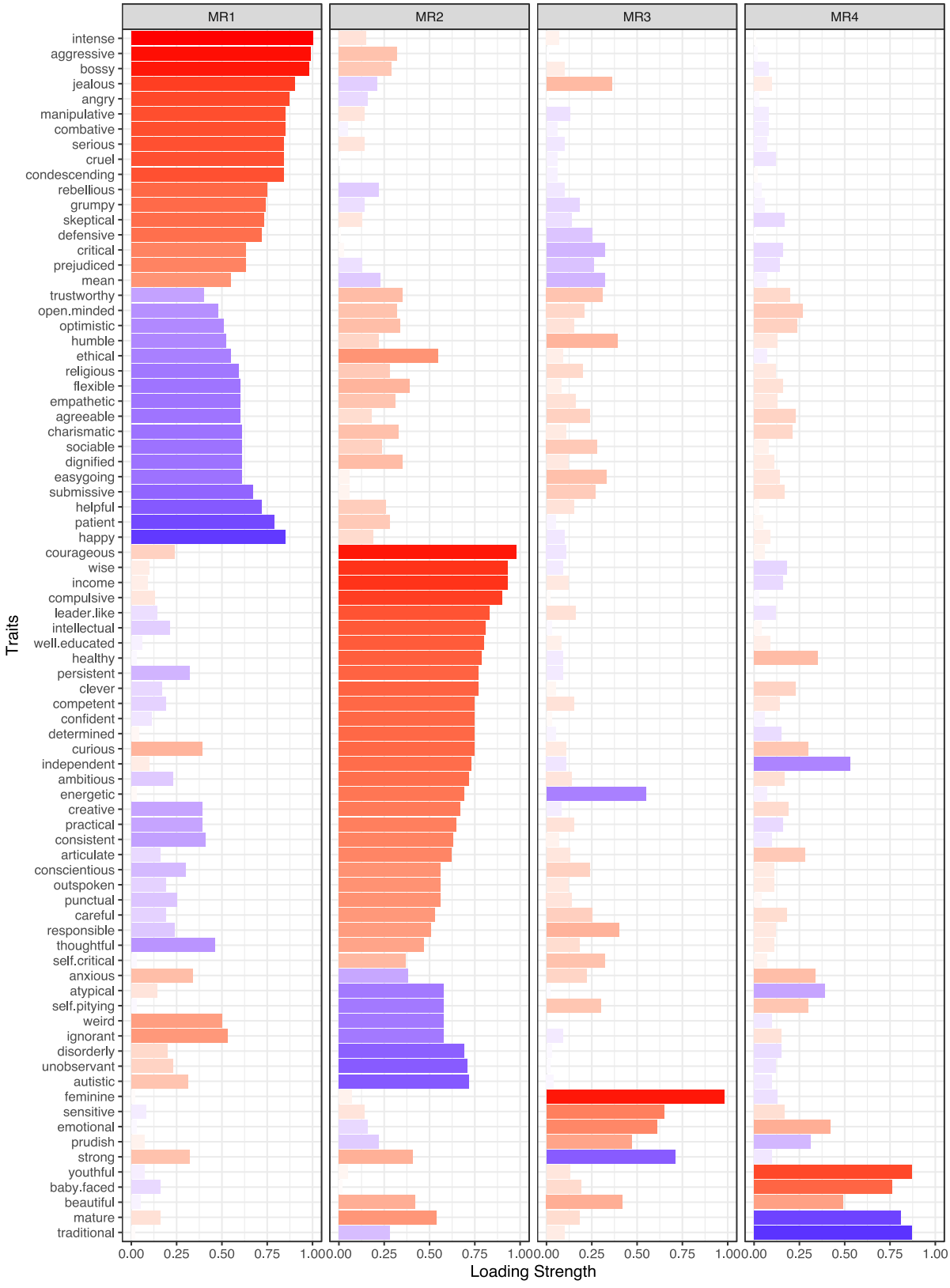
D


MR1
MR2
MR3
MR4
Traits
bossy
angry
manipulative
defensive
intense
grumpy
condescending
mean
aggressive
combative
prejudiced
jealous
cruel
critical
outspoken
rebellious
skeptical
serious
thoughtful
helpful
agreeable
empathetic
submissive
dignified
religious
patient
humble
leader.like
persistent
income
competent
healthy
wise
flexible
intellectual
clever
determined
articulate
ambitious
practical
curious
well.educated
optimistic
confident
beautiful
creative
consistent
charismatic
courageous
punctual
compulsive
energetic
sociable
conscientious
responsible
ethical
open.minded
independent
happy
careful
trustworthy
easygoing
self.critical
self.pitying
atypical
unobservant
anxious
ignorant
disorderly
autistic
weird
feminine
sensitive
emotional
prudish
strong
youthful
baby.faced
mature
traditional
0.00
0.25
0.50
0.75
1.00
Loading Strength

E


MR1
MR2
MR3
MR4
Traits
jealous
mean
angry
combative
persistent
critical
aggressive
defensive
outspoken
skeptical
grumpy
prejudiced
determined
compulsive
sensitive
patient
energetic
curious
ambitious
intellectual
income
healthy
practical
optimistic
independent
competent
clever
beautiful
open.minded
creative
confident
wise
happy
sociable
consistent
ethical
charismatic
thoughtful
responsible
flexible
courageous
bossy
manipulative
leader.like
articulate
well.educated
conscientious
condescending
punctual
self.critical
weird
unobservant
atypical
ignorant
disorderly
autistic
feminine
emotional
self.pitying
agreeable
baby.faced
religious
humble
submissive
helpful
easygoing
careful
empathetic
youthful
dignified
trustworthy
prudish
rebellious
strong
cruel
mature
traditional
anxious
serious
intense
0.00
0.25
0.50
0.75
1.00
Loading Strength

F


MR1
MR2
MR3
MR4
Traits
intense
aggressive
bossy
jealous
angry
manipulative
combative
serious
cruel
condescending
rebellious
grumpy
skeptical
defensive
critical
prejudiced
mean
trustworthy
open.minded
optimistic
humble
ethical
religious
flexible
empathetic
agreeable
charismatic
sociable
dignified
easygoing
submissive
helpful
patient
happy
courageous
wise
income
compulsive
leader.like
intellectual
well.educated
healthy
persistent
clever
competent
confident
determined
curious
independent
ambitious
energetic
creative
practical
consistent
articulate
conscientious
outspoken
punctual
careful
responsible
thoughtful
self.critical
anxious
atypical
self.pitying
weird
ignorant
disorderly
unobservant
autistic
feminine
sensitive
emotional
prudish
strong
youthful
baby.faced
beautiful
mature
traditional
0.00
0.25
0.50
0.75
1.00
Loading Strength

G

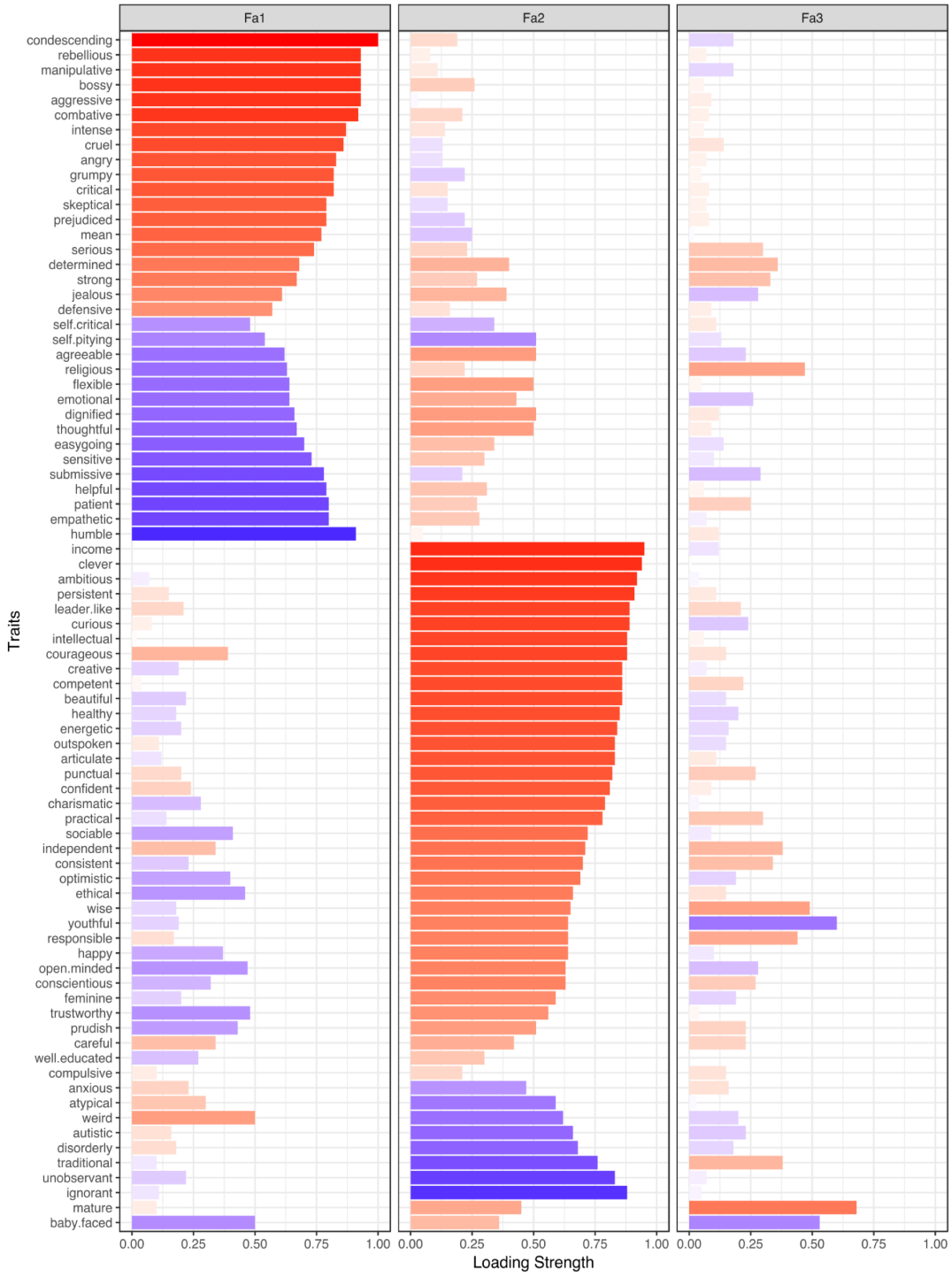


Fig. S9. Standardized factor loadings of 80 trait attributions across seven samples.
The seven panels plot results for samples from North America (**A**), Latvia (**B**), Peru (**C**), the Philippines (**D**), India (**E**), Kenya (**F**), and Gaza (**G**). Each column plots the strength of the factor loadings across the 80 traits. The color of the bar indicates the sign of the loading (red: positive; blue: negative); the length and saturation of the bar indicate the magnitude of the loading.

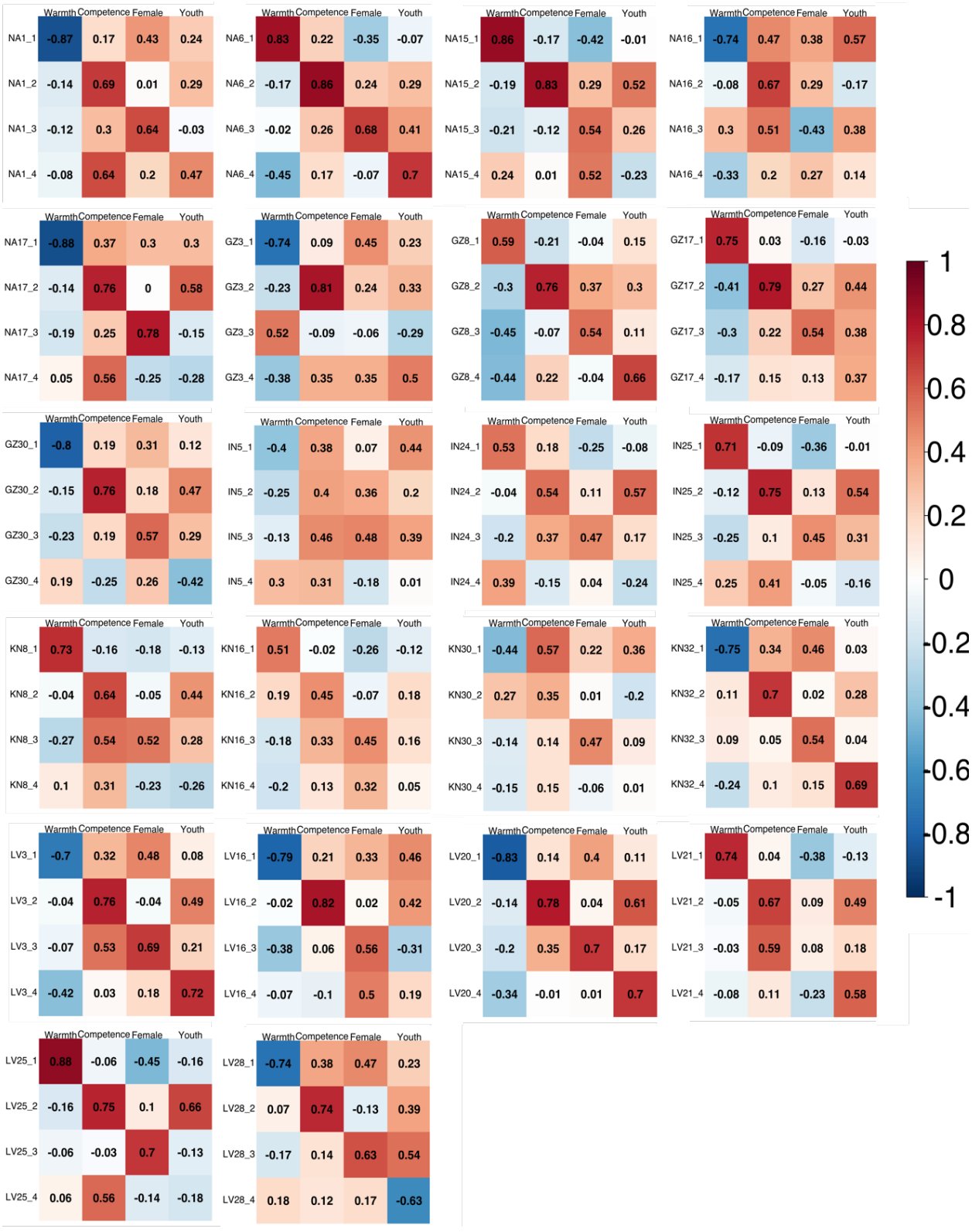


Fig. S10. Comparing the four dimensions obtained from aggregated and individual data.

The 22 panels plot the Tucker indices of factor congruence between the four dimensions found in aggregate-level data in Study 1 (columns) and those found in individual-level data in Study 2 (rows). Row names of each panel indicate the location (NA for North America, LV for Latvia, GZ for Gaza, KN for Kenya, IN for India) and ID of the participant, and the four dimensions uncovered from the individual's data (following the underscore). The numbers report the tucker indices. The color scale shows the sign and strength of the indices.

TRAITS	DEFINITIONS
ABUSIVE	A person who is extremely offensive and insulting
AFFECTIONATE	A person who is comfortable showing his/her love, warmth, and kindness
AGGRESSIVE	A person who pursues his/her aims and interests forcefully, sometimes with physical force
AGREEABLE	A person who is kind, cooperative, and sympathetic
AMBITIOUS	A person who has a strong desire and determination to succeed in their goals
ANGRY	A person who is usually angry
ANXIOUS	A person who stresses and worries about things
ARTICULATE	A person who speaks fluently and clearly, and who can express their ideas well
ATYPICAL	The structure, texture, shape or other aspects of the appearance of the face is unusual or rare
AUTISTIC	A person who has autism spectrum disorder—a developmental disorder characterized by troubles with social interaction and communication, and by restricted and repetitive behavior
BABY-FACED	A person who has facial features resembling a baby
BEAUTIFUL	A person who looks appealing and physically attractive
BOSSY	A person who likes giving people orders and wants things his/her own way
CAREFUL	A person who works and thinks in a cautious, thorough, or thoughtful way to avoid potential danger
CHARISMATIC	A person who is interesting and likeable because they have a charming personality
CLEVER	A person who is quick to understand and learn, and who can figure things out quickly
COMBATIVE	A person who likes to argue or pick a fight
COMPETENT	A person who is efficient and capable to do things in general
COMPULSIVE	A person who has to do things in a certain way and often checks and does things over and over again to make sure they are done exactly right
CONDESCENDING	A person who thinks he/she is better than others and puts other people down
CONFIDENT	A person who is sure about his/her own abilities, correctness, and successfulness
CONSCIENTIOUS	A person who does his/her work or duty thoroughly and responsibly
CONSERVATIVE	A person who sticks to traditional values, especially in politics or religion, and who does not like new ideas or changes
CONSISTENT	A person who behaves or responds in the same way over time; reliable
COURAGEOUS	A person who is not afraid to do the right thing, even if it is dangerous to them
CREATIVE	A person who has good imagination or original ideas
CRIMINAL	A person who looks like they could commit a crime
CRITICAL	A person who judges others harshly, and often makes disapproving comments
CRUEL	A person who willfully causes pain or suffering to other people or to animals, and feels no concern about it
CURIOUS	A person who is eager to learn about or experience new things
DEFENSIVE	A person who is easily offended and always guards themselves against criticism
DETERMINED	A person who is able to make firm decisions and is resolved not to change them
DIGNIFIED	A person who is polite and composed, and always shows good and respected manners
DISORDERLY	A person who is untidy and not organized

EASYGOING	A person who is relaxed, tolerant, and not prone to rigid rules or bouts of temper
EMOTIONAL	A person who shows his/her feelings and laughs and cries easily
EMPATHETIC	A person who is able to understand and share the feelings of others
ENERGETIC	A person who is very active and full of energy
ENTHUSIASTIC	A person who is filled with eager enjoyment and interest
ETHICAL	A person who is careful to do things that are morally right to do
FEMININE	A person whose facial appearance looks like a woman
FLEXIBLE	A person who is ready and able to change so as to adapt to different circumstances
GRUMPY	A person who is bad-tempered and always complaining
HAPPY	A person who is usually cheerful
HEALTHY	A person who is in good health
HELPFUL	A person who gives help when others are in need
HOMOSEXUAL	A person who is sexually attracted to people of his/her own sex
HUMBLE	A person who is modest and does not boast
IDIOT	A person who is stupid
IGNORANT	A person who doesn't know anything, and is also usually unaware of that
INCOME	A person's income level
INDEPENDENT	A person who is able to think and act without being influenced by others
INTELLECTUAL	A person who thinks a lot about the deeper meaning of things and likes to analyze things
INTENSE	A person who is very serious and expresses strong feelings
JEALOUS	A person who feels resentment about what other people have
LEADER-LIKE	A person who can take charge and help a group accomplish a goal
LOSER	A person who fails frequently or is generally unsuccessful in life
MANIPULATIVE	A person who likes to control people in order to meet his/her own needs
MATURE	A person who thinks and behaves like a responsible adult
MEAN	A person who is unkind, inconsiderate, and doesn't share things
NATURAL	A person who is relaxed and spontaneous
NOSEY	A person who is overly curious about other people's business
OPEN-MINDED	A person who is willing to try new things or to hear and consider new ideas
OPTIMISTIC	A person who is hopeful and confident about the future
OUTSPOKEN	A person who is frank in stating his/her opinions especially if they are critical or controversial
PASSIVE	A person who allows things to happen or accepts what others do, without resistance or trying to change anything
PATIENT	A person who is able to accept or tolerate delays or problems and is very relaxed about getting things done
PERSISTENT	A person who is able to continue in a course of action in spite of difficulty or opposition
PRACTICAL	A person who is sensible and realistic in dealing with a situation or problem
PREJUDICED	A person who holds biased judgments about other people; bigoted
PRUDISH	A person who is overly proper and cannot stand hearing any sexual reference
PUNCTUAL	A person who is always on time

REASONABLE	A person who makes sense and whose opinions most people would agree with
REBELLIOUS	A person who resists authority, control, or convention and wants to have their own way
RELIGIOUS	A person who practices religion and believes in their faith
RESERVED	A person who tends not to show their emotions or opinions and is quiet
RESPONSIBLE	A person who accepts the consequences of his or her own actions and decisions
SARCASTIC	A person who likes using irony in order to mock others
SELF-CRITICAL	A person who holds himself/herself responsible for any failures, always questioning if they did the right thing or not
SELF-PITYING	A person who feels sorry for themselves
SENSITIVE	A person who is aware of or careful about others' attitudes, feelings, or circumstances
SERIOUS	A person who shows deep thoughts and who doesn't smile or laugh easily
SHALLOW	A person who is concerned only about silly or inconsequential things; superficial
SINCERE	A person who says what he/she genuinely feels or believes
SKEPTICAL	A person who questions things and is not easily convinced
SOCIABLE	A person who is friendly and enjoys talking and engaging in activities with other people
STRICT	A person who follows rules exactly, and expects others to follow rules exactly
STRONG	A person who is physically vigorous and is able to exert great bodily or muscular power
SUBMISSIVE	A person who shows a willingness to be controlled by others or conforms to the authority or will of others
THOUGHTFUL	A person who is considerate of others' needs
THRIFTY	A person who uses money and other resources carefully and not wastefully
TRADITIONAL	A person who likes to do things the way they have always been done and accepted in the past
TRUSTFUL	A person who tends to trust other people easily (note: this is different from being trustworthy)
TRUSTWORTHY	A person who can be relied on as honest and truthful
UNOBSERVANT	A person who does not notice things
WEIRD	A person who does strange or bizarre things
WELL-EDUCATED	A person who has completed a high level of education, such as bachelor's, master's and doctorate degrees
WHITE	A person whose face looks like they are Caucasian
WISE	A person who has mature experience, knowledge, and good judgments
YOUTHFUL	A person who looks young

Table S1. Definitions of traits.

Definitions of traits were obtained from Google dictionary, with necessary modifications to make the definition easy to understand and fit the context of describing a person.

A

Traits from our set [traits in (*8*)]	Trustworthy	Dominant
Sociable [Sociable]	**0.89**	0.14
Weird [Weird]	**-0.88**	0.13
Beautiful [Attractive]	**0.86**	0.03
Confident [Confident]	**0.85**	-0.53
Responsible [Responsible]	**0.82**	0.12
Trustworthy [Trustworthy]	**0.77**	0.38
Wise [Wise]	**0.70**	-0.06
Thoughtful [Caring]	**0.64**	0.55
Happy [Unhappy]	**0.54**	0.45
Submissive [Dominant]	-0.18	**1.00**
Aggressive [Aggressive]	-0.13	**-0.90**
Mean [Mean]	-0.22	**-0.86**
Emotional [Emotionally stable]	0.48	**0.54**

B

Traits from our set [traits in (*35*)]	Approachability	Youthful/Attractive	Dominant
Wise [Intelligent]	**0.92**	-0.37	0.02
Trustworthy [Trustworthy]	**0.80**	0.20	0.24
Agreeable [Approachable]	**0.68**	0.20	0.43
Confident [Confident]	**0.63**	0.13	-0.63
Happy [No Smile-Big Smile]	**0.61**	0.21	0.26
Beautiful [Attractive]	**0.60**	0.54	-0.23
Feminine [Feminine]	**0.31**	0.28	0.20
Youthful [Youthful]	-0.11	**0.98**	0.12
Baby-faced [Baby-faced]	-0.09	**0.82**	0.31
Healthy [Healthy]	0.52	**0.67**	-0.25
White [Pallid-Tanned]	0.16	**0.27**	0.05
Submissive [Dominant]	0.05	0.21	**0.88**
Aggressive [Aggressive]	-0.38	-0.12	**-0.79**

Table S2. EFA loadings for subsets of data corresponding to existing theories.

(**A**) EFA on the 13 traits from our set (first column) that are the same or most similar to those in (*8*) [in brackets]. Two factors—the optimal number of factors as indicated by both Kaiser's Rule and Cattell's Scree Test—were extracted and rotated with oblimin. Each column lists standardized factor loadings across the 13 traits. The largest absolute loading across factors for each trait is highlighted in bold. (**B**) EFA on the 13 traits from our set (first column) that are the same or most similar to those in (*35*) [in brackets]. Three factors—the optimal number of factors as indicated by Kaiser's Rule, Cattell's Scree Test, and the optimal coordinates index—were extracted and rotated with oblimin. Each column lists standardized factor loadings across the 13 traits. The largest absolute loading across factors for each trait is highlighted in bold.

Traits	Warmth[reversed]	Competence	Female-stereotype	Youth-stereotype
Jealous	**-0.98**	-0.12	0.19	0.18
Critical	**-0.96**	0.18	0.00	0.11
Patient	**0.88**	0.02	0.18	0.08
Easygoing	**0.81**	-0.05	0.05	0.29
Agreeable	**0.75**	0.16	0.10	0.31
Mature	-0.06	**0.99**	0.13	-0.22
Practical	0.16	**0.88**	0.04	0.02
Leader-like	-0.16	**0.83**	-0.11	0.38
Weird	-0.07	**-0.67**	-0.11	-0.34
Idiot	-0.18	**-0.67**	-0.14	-0.34
Strong	-0.11	0.20	**-0.95**	-0.06
Prudish	-0.07	0.17	**0.91**	-0.15
Feminine	-0.07	0.06	**0.91**	0.11
Self-critical	-0.02	0.18	**0.76**	-0.06
Healthy	0.02	0.27	0.02	**0.81**
Youthful	0.11	-0.45	0.23	**0.77**
Charismatic	0.18	0.29	0.09	**0.69**
Sociable	0.22	0.33	0.11	**0.65**

Table S3. EFA loadings of the smallest trait set that produced the four dimensions.
Each column lists the standardized factor loadings across the 18 traits. The largest absolute loading across four factors for each trait is highlighted in bold. The four dimensions accounted for 88% of the common variance in this subset of data.